BMJ Open

The STI re-emergence in Catalonia (2017-2019): epidemic characterization, socio-epidemiological clustering approach, and HIV co-infection associated factors.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-052817
Article Type:	Original research
Date Submitted by the Author:	08-May-2021
Complete List of Authors:	Sentís, Alexis; CEEISCAT; Epiconcept SAS, Epidemiology Montoro-Fernandez, Marcos; CEEISCAT; IGTP Lopez-Corbeto, E; CEEISCAT; CIBERESP Egea-Cortés, Laia; CEEISCAT; IGTP Nomah, DK ; CEEISCAT; Autonomous University of Barcelona, Department of Paediatrics, Obstetrics and Gynecology and Preventive Medicine Díaz, Yesika; CEEISCAT; IGTP Garcia de Olalla, Patricia; Public Health Agency of Barcelona, Epidemiology; CIBERESP Mercuriali, Lilas; Public Health Agency of Barcelona, Epidemiology Borrell, Núria; Public Health Agency of Catalonia, Epidemiological Surveillance and Response to Public Health Emergencies Service in Tarragona Reyes-Urueña, Juliana; CEEISCAT; CIBERESP Casabona, Jordi; ASPC, CEEISCAT; CIBERESP
Keywords:	Epidemiology < INFECTIOUS DISEASES, HIV & AIDS < INFECTIOUS DISEASES, SEXUAL MEDICINE, Infection control < INFECTIOUS DISEASES, Public health < INFECTIOUS DISEASES, PREVENTIVE MEDICINE

The STI re-emergence in Catalonia (2017-2019): epidemic characterization, socio-epidemiological clustering approach, and HIV co-infection associated factors.

Alexis Sentís[1,2,3], Marcos Montoro-Fernandez[1,4], Evelin Lopez-Corbeto[1,4,5], Laia Egea-Cortés[1,4], Daniel Kwakye Nomah[1,6], Yesika Díaz[1,4], Patrícia Garcia de Olalla[5,7], Lilas Mercuriali[7], Núria Borrell[8], Juliana Reyes-Urueña[1,4,5], Jordi Casabona[1,4,5,6], and the Catalan HIV and STI surveillance group*

[1]Centre of epidemiological studies on sexually transmitted infections and AIDS of Catalonia (CEEISCAT). Department of Health. Generalitat of Catalonia. Badalona. Spain.

[2]Pompeu Fabra University (UPF), Barcelona, Spain.

[3]Epiconcept, Epidemiology department, Paris, France.

[4]Fundació Institut d'Investigació Germans Trias i Pujol (IGTP), Badalona, Spain.

[5]Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain.

[6]Department of Paediatrics, Obstetrics and Gynecology and Preventive Medicine, Universitat Autònoma de Barcelona, Badalona, Spain.

[7]Epidemiology Service. Public Health Agency of Barcelona, Barcelona, Spain.

[8]Epidemiological Surveillance and Response to Public Health Emergencies Service in Tarragona, Agency of Public Health of Catalonia, Generalitat of Catalonia, Tarragona, Spain

*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 Mercuriali, 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 Martinez, 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 Mendioroz, 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).

Corresponding author: Alexis Sentís, e-mail address: alexissentis@gmail.com, address: Fundació Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol (IGTP), Edifici Muntanya, Carretera de Can Ruti, Camí de les Escoles s/n, 08916 Badalona (Spain).

Word counts for the abstract: 300

Word counts for the main text: 3215

Number of references: 30

Number of figures: 1

Number of tables: 3

Supplementary material: 4 tables.

Key words:

Epidemiology, infection control, public health, HIV & AIDS, sexual medicine, preventive medicine

Abstract

Objectives: The objectives of this study were to describe the epidemiological characteristics of the STI cases, to identify STI-HIV coinfection associated factors, and to identify and characterize STI socio-epidemiological clusters in Catalonia between 2017-2019.

Design: A population-based retrospective cohort study.

Participants: All STI notified confirmed cases of-syphilis, gonorrhea, chlamydia, LGV and HIV-, between 2017-2019, to the Catalan HIV/STI Registry of Catalonia.

Primary and secondary outcomes: We performed a descriptive analysis of STI confirmed cases using all notified STI and HIV cases. Factors associated with HIV coinfection were determined using logistic regression. We identified and characterized STI socio-epidemiological clusters by basic health area (ABS) using K-means clustering methodology.

Results: The STI-cases were doubled, primarily due to the increase in chlamydia and gonorrhea in women and people younger than 30 years of age, 11% were reinfections, and 6% coinfected with HIV. Syphilis and LGV occurred more frequently in men who have sex with men (MSM), gonorrhea in heterosexual people, and chlamydia in heterosexual women. Men, aged 30-60, living in urban and less deprived ABS, and having multiple STI-episodes were associated with an increased risk of HIV coinfection. When comparing the distribution of proportions of socio-epidemiological characteristics' in the overall STI-cases with those within the three clusters of ABS identified(A, B, and C), we found in A) similar distribution-values; B) higher proportion of chlamydia, women, younger people, heterosexuals, and people living in rural and more deprived areas; and

C) higher incidence rates for all STI, higher proportion of MSM, multiple episodes, HIV coinfection, and higher proportions of people living in urban and less deprived areas.

Conclusions: STI increased dramatically in Catalonia mostly in women and young people. We identified and characterized three socio-epidemiological clusters, which, along with the associated HIV confection factors, provides a characterization of key populations at a small area level.

Strengths and limitations of this study

- We found that STI increases dramatically, not only among men who have sex with men (MSM) but also among heterosexual women and young adults.
- Our study shows that men, aged 30-60, living in urban and less deprived areas, and having multiple STI episodes were associated with an increased risk of HIV coinfection.
- To our knowledge, for the first-time k-means clustering methodology has been used to identify and characterize different STI socio-epidemiological clusters at a small health area level.
- MSM, heterosexual women and young adults need to be considered a priority for the STI and HIV preventive strategies taking into account the structural determinants also identified as crucial in our analysis.
- A limitation of the study is the high proportion of missing values in relevant variables, such are education level, sexual preference, or country of birth. Nonetheless, the descriptive analysis in some related reports or studies shows very similar results for these variables.

Introduction

The epidemic of sexually transmitted infections (STI) is a major public health concern in high-income-, middle-, and low-income countries. Daily, over 1 million people acquire STI worldwide [1,2]. When STI are not detected early and treated properly, it can make the infected individual prone to a magnitude of related complications including HIV acquisition [3], long-term disabilities, infertility, adverse pregnancy outcomes and death [1,2]. Furthermore, people diagnosed with STI suffer from varying levels of stigma, shame, stereotyping and have been subjected to gender-based violence [4]. In Europe (EU/EEA) the incidence rates of gonorrhoea, syphilis and Lymphogranuloma venereum (LGV) have been increased 50%, 36% and 69%, respectively, from 2014 to 2018 [5]. Similarly, in Spain, in 2017 alone, there were 23,975 cases of gonorrhoea, syphilis, chlamydia and LGV [6], which represents more than a 10 fold increase, from those reported in 2000 [7]. Catalonia recorded the highest incidence across the country in all these infections, with a rise of 37% between 2018 and 2019 [8]. Rates were highest among men who have sex with men (MSM) and young adults, mostly women, who recently showed a proportionally higher increase [6,8]. The surge in the incidence rates of STI can be explained by improvements on the surveillance systems, the introduction of new more sensitive diagnostic tools, variations in sexual behaviours, sociocultural changes, tourism or globalization, among others factors [5,9,10].

STI and HIV are overlapping epidemics, which a part from biological synergies, are mostly driven by socioeconomic and other contextual factors acting as syndemics. Therefore it is crucial, that the different actors working in their prevention and control collaborate towards an integrated STI/HIV/behavioural surveillance in order to identify and characterize groups of greater risk of STI/HIV acquisition [9–11]. People affected by an STI are more likely to be at risk to acquire HIV and HIV positive people are more vulnerable to STI [12,13]. Some studies have identified the social determinants of health, discrimination, and inequities as main factors associated with the appearance of STI spatiotemporal clustering of cases [14–16]. The identification and characterization of STI socio-epidemiological clusters and their association with HIV coinfection is imperative to strengthen STI/HIV integrated surveillance and increase its sensitivity, timeliness and representativeness, but also to generate strategic information, to tailor public health strategies in order to tackle a growing hidden epidemic. The objectives of this study were to describe the epidemiological characteristics of the STI cases, to identify STI-HIV coinfection associated factors, and to identify and characterize STI socio-epidemiological clusters in Catalonia from 2017 to 2019.

Methods

Study design, participants and surveillance systems

A population-based retrospective cohort study was conducted of the STI notified confirmed cases of -syphilis, gonorrhoea, chlamydia, LGV and HIV-, between January 1, 2017 and December 31, 2019, to the Catalan HIV/STI Registry of Catalonia [17], which use data from the Epidemiological Repository of Catalonia (REC) reported by means of a standardized notification, as well as complementary epidemiological questionnaires (see supplementary material, Table S1). Case definitions follows the European standardized case definitions stablished by ECDC [18]. All individuals who had experienced one or more episodes of STI during the study period were also linked, through the Spanish health system personal identification code (CIP), to the archive of the HIV/STI Registry of Catalonia in order to identified coinfections, either before or after the STI episode. We used a basic health area (ABS) deprivation index categorized in quintiles, first quintile for the ABS with lower deprivation index [19]. The clinical variables were reinfections, multiple STI episodes (total number of episodes due to any STI that had the same person during the study period), and coinfection with HIV. Reinfection was defined as more than one episode of the same specific STI during all the study follow-up, but defined differently for each STI, depending on the number of days between successive episodes after the first infection in the same individual. In Table S2 (supplemental material) we described all the used variables.

Statistical analyses

We performed a descriptive analysis of all the epidemiological, clinical, and geographical variables for the new cases of STI - syphilis, gonorrhoea, chlamydia and LGV infections. For quantitative variables, we used measures of central tendency and dispersion (mean, standard deviation, median, and interquartile range). For qualitative variables, we calculated absolute frequencies and percentages. For the descriptive analysis, we calculated annual incidence rates per 100,000 inhabitants for the overall and each specific STI and Catalan health region based on census information from the Statistical Institute of Catalonia (IDESCAT). For the K-means clustering analysis we calculated, for the overall and each specific STI through all the study period (2017-19), the incidence rate per 1,000 inhabitants by each ABS. An ABS is a geographical area where a specific primary basic health team is in charge of visiting its population and covers territories with a population of between 5,000-25,000 people approximately.

We assessed the risk factor of HIV coinfections among persons diagnosed with STI using multivariable logistic regression models. Persons with more than one STI episode were counted just one time (first episode), and the successive episodes in the same person were grouped in a variable, which counts the number of episodes, and was included in the models. Potential risk factors that showed a statistically significant association with HIV diagnosis in the univariate analysis were included in multivariable logistic regression models. Variables such as sexual preference and education level were excluded in the models because missing values were higher than 50% (Supplemental material, table S3).

We performed a cluster analysis using K-means clustering methodology which is a machine learning technique that identifies groups by specific unit of analysis (in our case the ABS) based on similarities in characteristics (variable values and categorical distribution among the ABS) [20]. The following variables were chosen and used by the afore-mentioned methods to identify the socio-epidemiological clusters of ABS between 2017-19; incidence rate by each STI, percentage of women, percentage of people with HIV coinfection, and median age among the all STI cases in each ABS, and deprivation index in each ABS. Finally, we performed a descriptive analysis of each cluster.

Odds ratios (OR) and its 95% confidence intervals (CI) were estimated. All analyses were performed using R version 3.6.1.

Ethics approval statement

Data from mandatory notifiable disease in REC and the rest of aggregated variables used in the study were handled according to the international recommendations [21], the Helsinki Declaration revised by the World Medical Organization in Fortaleza in 2013, and to Spanish Law 3/2018 on Data protection and Public Health 33/2011. Patient information was anonymised and de-identified prior to analysis and therefore no informed consent was required.

Patient and public involvement

Patients were not directly involved in this study; only data coming from notifiable disease surveillance systems were used.

Results

Between 2017 and 2019, there were 42,283 cases of STIs in Catalonia (an increase of 51% from 2017 to 2019): 21,202 cases of chlamydia, 13,362 of gonorrhea, 6,975 of syphilis, and 744 of LGV. During the study period, the highest STI incidence rate was for chlamydia that had also the highest increase,188%, the lowest increase was for syphilis, which remained with similar values. By health regions, Barcelona had the highest incidence rate throughout the study period and Alt Pirineu i Aran the lowest. Urban ABS presented higher rates than rural ones, 179.9 vs 26.3 per 100,000 inhabitants respectively in 2019 (see supplemental material, table S3).

From the total number of STI cases during the study period 40% were in women. Proportionally, the STI episodes in men were significantly higher than in women, for gonorrhoea, syphilis, and LGV, but less frequent for chlamydia (80%, 87%, 99%, and 38% were in men, respectively). Among those cases with sexual preference information available (35%), 95% were WSM. In men, 51% were MSM. Among STI cases, 77% were in people younger than 40 years. Chlamydia was the STI with the highest proportion of cases among people younger than 30 years (65%) and syphilis was the most prevalent among people of 40 years and older (45%). Regarding socioeconomical status, from the overall cases, the highest proportion (24%) was in the first quintile -lower deprivation index-, whereas the lowest (18%) was in the fifth quintile. People that had chlamydia and gonorrhea episodes, showed higher deprivation indexes (higher proportions in the 5th quintile) than those who had syphilis and LGV. Among all the reported cases 11% were reinfections: Gonorrhea had the highest proportion (15.70%), whereas chlamydia the lowest (7 %). The STI episodes in HIV-positive counted 6% from the overall, however, with higher proportion in syphilis and LGV (13% and 25%, respectively) and the lowest in chlamydia (2%) (see table 1). Despite high missing values for country of birth (57%)

and education level (76%), among those episodes with available information, 72% were born in Spain and 85% had secondary or higher education (see supplemental material, table S4).

In a multivariate analysis: being a man (adjusted Odds Ratio (ORa) 23.69, 95% confidence interval (CI): 16.67 - 35.13 vs. women), being over 20 years of age (ORa 18.58 (CI: 8.56 - 52.13) for the people aged between 30 to 39 years old vs younger than 20 years old), having been diagnosed with more than one episode of STI between 2017-19 (ORa 5.96 (CI: 4.26 - 8.24 for those who presented between 5 and 7 episodes vs. those with one episode), and living in urban ABS (ORa 1.32 (IC: 1.04 - 1.69)), were associated with an increased risk of HIV coinfection. Belonging to the fifth quintile -higher deprivation index- (ORa 0.6 (IC: 0.5 - 0.72) for those in the fifth quintile vs. those in the first quintile) was associated with lower risk of HIV coinfection (see table 2).

The distribution of proportions of people's socio-epidemiological characteristics among cases within Cluster A of ABS were similar than that presented for all the cases. When comparing this distribution within STI cases in Cluster B of ABS with that from all the cases, we observed that Cluster B: i) had predominantly younger people (median age of 26 vs 29 years), who were living in ABS with higher deprivation index (44.87 vs 39.82); ii) a higher proportion were women (53% vs 42%), more percentage declare being heterosexual -women or men- (approximately 20% higher), more proportion of chlamydia episodes (62% vs 55%) and a relatively higher proportion lived in rural ABS (16% vs 11%); and iii) presented lower STI incidence rate by ABS (79.46 vs 97.37), lower proportion of multiple STI episodes (7.64% vs 11.64%) and HIV coinfection (2% vs 6%). When comparing the distribution of characteristics' proportion in all the STI cases with that observed within the cases within Cluster C of ABS, composed of only by eight ABS, we observed that Cluster C: i) had older people (median age of 34 vs 29 years old), living in ABS with lower deprivation index (25.62 vs 39.82), ii) reported a higher STI incidence rate by ABS (544.87 vs 97.37), were predominantly men (84.26% vs 58.11%), had a high proportion of MSM (approximately 20% higher), reported higher episodes of gonorrhoea, syphilis and LGV (33%, 24% and 5% vs 28%, 15% and 2%), showed a higher percentage of multiple STI episodes (24% vs 12%) and HIV coinfection (16% vs 6%), and all episodes were from people living in urban ABS (see table 3).

More than 60% of the episodes in cluster A occurred in ABS with high STI incidence rates (in quintiles; 4th and 5th quintile), in the cluster B almost 60% of the episodes were in ABS classified in the three lowest quintiles (1st to 3rd), and in the cluster C all the 4,359 episodes were in the highest quintile ABS of STI incidence rate (see table 3 and Figure 1).

Discussion

Between 2017 and 2019, the number of STIs has doubled in Catalonia primarily due to the increase in chlamydia and gonorrhoea in women and people younger than 30 years of age. A higher proportion of STI diagnoses occurred in people living in urban and less deprived areas. Among all the STI cases, 6% were reinfections by the same STI and 11% were coinfected with HIV. Despite a low response rate regarding sexual preference in the epidemiological questionnaire (35%), syphilis and LGV seemed to occur more frequently in MSM, gonorrhoea in HSW and WSM, and chlamydia in WSM. Men, aged 30-60, living in urban and less deprived areas and having multiple STI episodes were associated with an increased risk of HIV coinfection. When comparing the distribution of the proportions of socio-epidemiological characteristics' in the overall STI cases with those within the three clusters (A, B, and C) of ABS identified, we found in A) similar distribution-values; B) higher proportion of chlamydia, women, younger people, heterosexuals, and people living in rural and more deprived areas; and C) higher incidence rates for all STI, higher proportion of MSM, multiples episodes, HIV coinfection, and higher proportions of people living in urban and less deprived areas.

After a long period of continuous reduction in the STI incidence in western countries, which coincided with the beginning and hardest times of the HIV epidemic, from 80's to 2010 approximately, many countries including the USA and European countries, have reported a re-emergence of the STI [22]. From our study, we can confirm that from 2017 to 2019, the total number of STI cases doubled in Catalonia, increasing in higher proportions for chlamydia and gonorrhoea, and affecting more frequently Barcelona health region and urban areas. MSM and more recently, young adults, mostly women, are the main population groups at risk [8,22]. Besides, in western countries, although this dynamics could potentially change, STIs seem to be affecting in higher proportion people with favourable socioeconomic status and educational levels [23]. Chlamydia has been reported more frequently in WSM and syphilis, gonorrhoea and LGV in heterosexual men (MSW) and MSM. The rise of STI cases have been partially attributed to surveillance systems strengthening and the introduction of better diagnostic tools over last year's [9,10]. Other factors that can be contributing to the raising STI epidemic, but that has been described mostly in MSM, are the use of Pre-Exposure Prophylaxis (PrEP), recreational drugs for sex, substance abuse, alcohol and prevalent use of internet and other enhanced technologies to find sexual partners [24–26]. In the present study, we found different epidemiological characteristics in each STI. Chlamydia was more common in young women, and in MSW and WSM. Regarding syphilis and LGV, were more frequent in men, more specifically among MSM, in people living in more deprived areas, and both infections showed higher percentages of reinfections and

HIV coinfections. Most of the STI cases were found in people born in Spain with secondary or higher education. But this should be interpreted with caution because of the high percentages of missing values for sexual preference and education level.

It has been described how HIV and STI are synergic infections and need be seen as a syndemic [11]. WHO and other public health agencies have pointed out the importance to integrate STI, HIV, and even blood borne diseases and behavioural surveillance in order to improve the identification and characterization of key populations [9–11]. Sociodemographic and socioeconomic related risk factors seem to be more associated than individual behaviours with STI acquisition, particularly in women from disadvantaged groups [27,28]. Also consistent with previous data [22], we found that men, aged 30-60, living in less deprived and rural areas, and having multiple episodes of STI were associated with an increased risk of HIV coinfection.

Last years, the K-means clustering methodology has proven its potential in classifying and grouping health related outputs in different study fields. For instance, in Bipolar disorder, it has been used [29] to obtain a cluster-based classification of severity using several variables with heterogeneous origins; socio-demographic, clinical, cognitive, vital signs, and lab analysis among others. More recently, its potential to monitor and group by magnitude (higher-medium-lower) SARS-CoV-2 prevalence and trend at a regional level in Italy has been described [30]. These classifications or "clusters of characteristics" identification may be useful, in each specific context, to better detect and characterize different case profiles by site or geographical area which ultimately could lead to better-designed interventions that would allow improvements in health results. In the present study, by using k-means clustering, to our knowledge for the first time in our setting, we identified three different STI socio-epidemiological clusters of ABS with the already mentioned specific characteristics.

A limitation of the study is the high proportion of missing values in relevant variables from the epidemiological questionnaires, such are education level, sexual preference, or country of birth, this fact could be biasing our results or not reflecting the real situation when describing the key population affected by STI. Despite this matter, our results were similar than those shown in international public health reports or studies [5,13,23]. A strength of our study is that we include an ecological variable of socio-economic status which bring a relevant information to describe the groups at more risk of being affected by STI. We believe that the most valuable input of our study is that it shows the utility of complementing the "traditional" epidemiological analysis performed by public health authorities by using new methodologies, such are machine learning techniques, to combine variables from heterogeneous sources. By using it, may allow to identify and characterize the target key populations to design more efficient measures to enhance STI/HIV prevention and control at a small health area level.

We conclude that, consistently with other European countries, STI increased dramatically from 2017 to 2019 in Catalonia, and it is a growing and hidden public health problem, above all in women and people younger than 30 years of age. STI epidemics is not only an issue of the health sector alone, it is a wide spectrum of the development agenda. While the HIV trend to decrease, mainly because the wider and earlier use of ARV, STI increases dramatically, not only among MSM but also among heterosexual women and young adults. These populations need to be considered a priority for the preventive strategies of STI taking into account the structural determinants also identified as crucial in our analysis. Young women living more in rural and deprived areas seemed more likely to be affected by chlamydia. MSM living in urban and less deprived areas showed, more frequently than other population groups, higher STI rates, more multiple STI episodes and higher percentages of HIV coinfection. Monitoring the epidemics of

STI in accordance with determinants of health and localized intervention programmes would have a paramount importance rather than using the national prevalence as the key monitoring variable.

Acknowledgements

We thank all the health 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 questionnaire.

Conflict of interest statement

All of the authors declare that they have no conflicts of interest.

Funding statement

The authors have not declared a specific funding.

Authors' contributions

AS conceptualized and designed the study. MM cleaned the database, MM, LE and YD performed the statistical and cluster analysis. AS, EL and DN reviewed scientific literature, AS, JR, JC and DN drafted the manuscript and AS, EL, JR and JC interpreted the results. All the authors collaborated in the critical review and approved the final manuscript.

Data sharing statement

No additional data available.

References

1 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**:548-562P. doi:10.2471/BLT.18.228486

2 WHO Department of Reproductive Health and Research. Report on global sexually transmitted infection surveillance. 2018. https://apps.who.int/iris/bitstream/handle/10665/277258/9789241565691-eng.pdf?ua=1

3 Zetola NM, Bernstein KT, Wong E, *et al.* Exploring the relationship between sexually transmitted diseases and HIV acquisition by using different study designs. *J Acquir Immune Defic Syndr* 2009;**50**:546–51. doi:10.1097/qai.0b013e318195bd2b

4 Amin A. Addressing gender inequalities to improve the sexual and reproductive health and wellbeing of women living with HIV. J. Int. AIDS Soc. 2015;**18**. doi:10.7448/IAS.18.6.20302

5 Annual Epidemiological Report for 2017: Sexually Transmitted Diseases. https://www.ecdc.europa.eu/en/publications-data/presentation-annual-epidemiological-report-2017-sexually-transmitted-diseases (accessed 7 Aug 2020).

6 Centro Nacional de Epidemiología. Unidad de vigilancia del VIH y conductas de riesgo. Vigilancia epidemiológica 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 P. 2019. https://www.mscbs.gob.es/ciudadanos/enfLesiones/enfTransmisibles/sida/vigilan

cia/Vigilancia_ITS_1995_2017_def.pdf

7 Área de vigilancia del VIH y conductas de riesgo. Vigilancia epidemiológica de las infecciones de transmisión sexual, 1995-2010. Madrid: Centro Nacional de Epidemiología. Published Online First: 2012.https://www.mscbs.gob.es/ciudadanos/enfLesiones/enfTransmisibles/sida/vigilancia/VigilanciaITS1995_2010.pdf

8 Centre d'Estudis Epidemiològics sobres les Infeccions de Transmissió Sexual i Sida de Catalunya (CEEISCAT). Vigilància epidemiològica de les Infeccions de Transmissió Sexual (ITS) a Catalunya Informes anual 2019. 2020. https://canalsalut.gencat.cat/web/.content/_A-Z/S/sida/enllasos/anual_ITS.pdf

9 Taylor MM, Wi TEC. Transforming and integrating STI surveillance to enhance global advocacy and investment in STI control. *J Int AIDS Soc* 2019;**22**:25361. doi:10.1002/jia2.25361

10 Mills S, Saidel T, Magnani R, *et al.* Surveillance and modelling of HIV, STI, and risk behaviours in concentrated HIV epidemics. In: *Sexually Transmitted Infections*. The Medical Society for the Study of Venereal Disease 2004. ii57–62. doi:10.1136/sti.2004.011916

11 Murti M, Wong J, Whelan M, *et al.* The need for integrated public health surveillance to address sexually transmitted and blood-borne syndemics. *Canada Commun Dis Rep* 2019;**45**:63–6. doi:10.14745/ccdr.v45i23a03

12 HIV/AIDS & STDs. https://www.cdc.gov/std/hiv/default.htm (accessed 22 Feb 2021).

13 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. doi:10.1136/bmjopen-2018-027245

14 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. doi:10.1017/S0950268813000745

15 Bernstein KT, Curriero FC, Jennings JM, *et al.* Defining core gonorrhea transmission utilizing spatial data. *Am J Epidemiol* 2004;**160**:51–8. doi:10.1093/aje/kwh178

16 Gesink D, Wang S, Norwood T, *et al.* Spatial epidemiology of the syphilis epidemic in Toronto, Canada. *Sex Transm Dis* 2014;**41**:637–48. doi:10.1097/OLQ.0000000000000196

17 SIVES 2015 - Sistema Integrat de Vigilància Epidemiològica de la SIDA/VIH/ITS a Catalunya. 2015. 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).

18 Jané i Checa M, Carmona i Parcerisa G. Definició de cas de les noves malalties de declaració individualitzades (MDI). 2016. https://scientiasalut.gencat.cat/bitstream/handle/11351/1955/definicio_cas_noves_malalties_declaracio_individualitzades_2016.pdf?sequence=1&isAllowed=y

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.http://observatorisalut.gencat.cat/ca/observatori-desigualtats-

salut/indicador_socioeconomic_2015/ (accessed 6 Aug 2020).

20 Steinley D. K-means clustering: A half-century synthesis. *Br J Math Stat Psychol* 2006;**59**:1–34. doi:10.1348/000711005X48266

21 Rose S. International Ethical Guidelines for Epidemiological Studies: By the Council for International Organizations of Medical Sciences (CIOMS). *Am J Epidemiol* 2009;**170**:1451–2. doi:10.1093/aje/kwp334

22 Annual Epidemiological Report for 2017: Sexually Transmitted Diseases. https://www.ecdc.europa.eu/en/publications-data/presentation-annual-epidemiological-report-2017-sexually-transmitted-diseases (accessed 16 Dec 2020).

23 Martí-Pastor M, García de Olalla P, Barberá M-J, *et al.* Epidemiology of infections by HIV, Syphilis, Gonorrhea and Lymphogranuloma Venereum in Barcelona City: a population-based incidence study. *BMC Public Health* 2015;**15**:1015. doi:10.1186/s12889-015-2344-7

24 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. doi:10.1093/eurpub/ckp091

25 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. 2019;**18**:1–11. doi:10.1177/2325958219867324

26 Jansen K, Steffen G, Potthoff A, *et al.* STI in times of PrEP: High prevalence of chlamydia, gonorrhea, and mycoplasma at different anatomic sites in men who

have sex with men in Germany. *BMC Infect Dis* 2020;**20**. doi:10.1186/s12879-020-4831-4

27 Hill A V., 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 Heal* 2019;**64**:319–26. doi:10.1016/j.jadohealth.2018.09.006

28 Jagadesh S. Mapping priority neighborhoods: A novel approach to cluster identication in HIV/AIDS population. doi:10.21203/rs.3.rs-17843/v1

29 de la Fuente-Tomas L, Arranz B, Safont G, *et al.* Classification of patients with bipolar disorder using k-means clustering. *PLoS One* 2019;**14**. doi:10.1371/journal.pone.0210314

30 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**:1–14. doi:10.3390/ijerph17155286

Figures (legends)

Figure 1. Map of the STI incidence rates (per 1,000 inhabitants) and STI socio-epidemiological clusters by basic health area (ABS), 2017-19: 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 in the manuscript as a bigger unit of analysis than ABS, in spite of this fact, in this figure we show the municipality of Barcelona in order to provide better visualization of Cluster C. From a total of 373 Catalan ABS five were excluded -Garraf rural, Polinyà-Sentmenat, Ribes-Olivella. Roquetes-Canyelles, Viladecans 3- from the K-means clustering analysis because their delimitations and populations changed during the study period.

Tables

Table 1. Distribution of epidemiological characteristics in STI cases of chlamydia, gonorrhoea, syphilis or lymphogranuloma venerum (LGV) in Catalonia, 2017–2019

	All STI (N=42,283)		Chlamydia (N=21,202)		Gonorrhoea (N=13,362)		Syphilis (N=6,975)		LGV N =744)	
	N	%	N	%	N	%	N	%	N	%
Sex										
Female	16,676	39.44	13,125	61.90	2,667	19.96	875	12.54	9	1.21
Male	25,607	60.56	8,077	38.10	10,695	80.04	6,100	87.46	735	98.79
Age group										
<20	4,438	10.50	3,311	15.62	984	7.36	137	1.96	6	0.81
20-29	17,691	41.84	10,707	50.50	5,361	40.12	1,462	20.96	161	21.64
30-39	11,102	26.26	4,454	21.01	4,116	30.80	2,242	32.14	290	38.98
40-49	6,092	14.41	2,087	9.84	2,032	15.21	1,757	25.19	216	29.03
50-59	2,037	4.82	530	2.50	658	4.92	789	11.31	60	8.06
>60	923	2.18	113	0.53	211	1.58	588	8.43	11	1.48
Sexual preference										
MSM[a]	3,270	7.73	785	3.70	1,321	9.89	993	14.24	171	22.98
MSW	3,149	7.45	1,863	8.79	1,040	7.78	243	3.48	3	0.40
WSW[b]	415	0.98	335	1.58	69	0.52	10	0.14	1	0.13
WSM	8,189	19.37	7,034	33.18	966	7.23	186	2.67	3	0.40
Missing men	19,188	45.38	5,429	25.61	8,334	62.37	4,864	69.73	561	75.40
Missing women	8,072	19.09	5,756	27.15	1,632	12.21	679	9.73	5	0.67
Deprivation index[c]										
1st quintile (lower deprivation)	10,271	24.29	5,185	24.46	3,040	22.75	1,757	25.19	289	38.84
2nd quintile	7,465	17.65	4,328	20.41	2,012	15.06	1,037	14.87	88	11.83
3rd quintile	4,859	11.49	2,763	13.03	1,332	9.97	716	10.27	48	6.45
4th quintile	5,703	13.49	3,217	15.17	1,578	11.81	827	11.86	81	10.89
5th quintile (higher deprivation)	7,689	18.18	4,319	20.37	2,211	16.55	1,079	15.47	80	10.75
Missing	6,296	14.89	1,390	6.56	3,189	23.87	1,559	22.35	158	21.24
Reinfection										
No	37,725	89.22	19,784	93.31	11,264	84.30	6,020	86.31	657	88.31
Yes	4,558	10.78	1,418	6.69	2,098	15.70	955	13.69	87	11.69
HIV coinfection										
No	39,840	94.22	20,735	97.80	12,465	93.29	6,082	87.20	558	75.00
Yes	2,443	5.78	467	2.20	897	6.71	893	12.80	186	25.00
Health region of residence										
Other health regions	7,068	16.72	4,094	19.31	1,796	13.44	1,142	16.37	36	4.84
Barcelona	35,215	83.28	17,108	80.69	11,566	86.56	5,833	83.63	708	95.16
ABS Urbanicity										
Rural	4,193	9.92	2,614	12.33	1,039	7.78	516	7.40	24	3.23
Urban	29,969	70.88	16,347	77.10	8,566	64.11	4,516	64.75	540	72.58
Missing	8,121	19.21	2,241	10.57	3,757	28.12	1,943	27.86	180	24.19

MSM[a]: men who have sex with men, bisexual men, transgender men, WSW[b]: women who have sex with women, bisexual women, transgender women, Deprivation index[c]: 1st quintile (31.52%), 2nd quintile (40.09%), 3rd quintile (46.27%), 4th quintile (53.98%), 5th quintile (100%)

Table 2. Associated factors to HIV coinfection in patients diagnosed of chlamydia, gonorrhoea, syphilis or lymphogranuloma venerum in Catalonia, 2017-19.

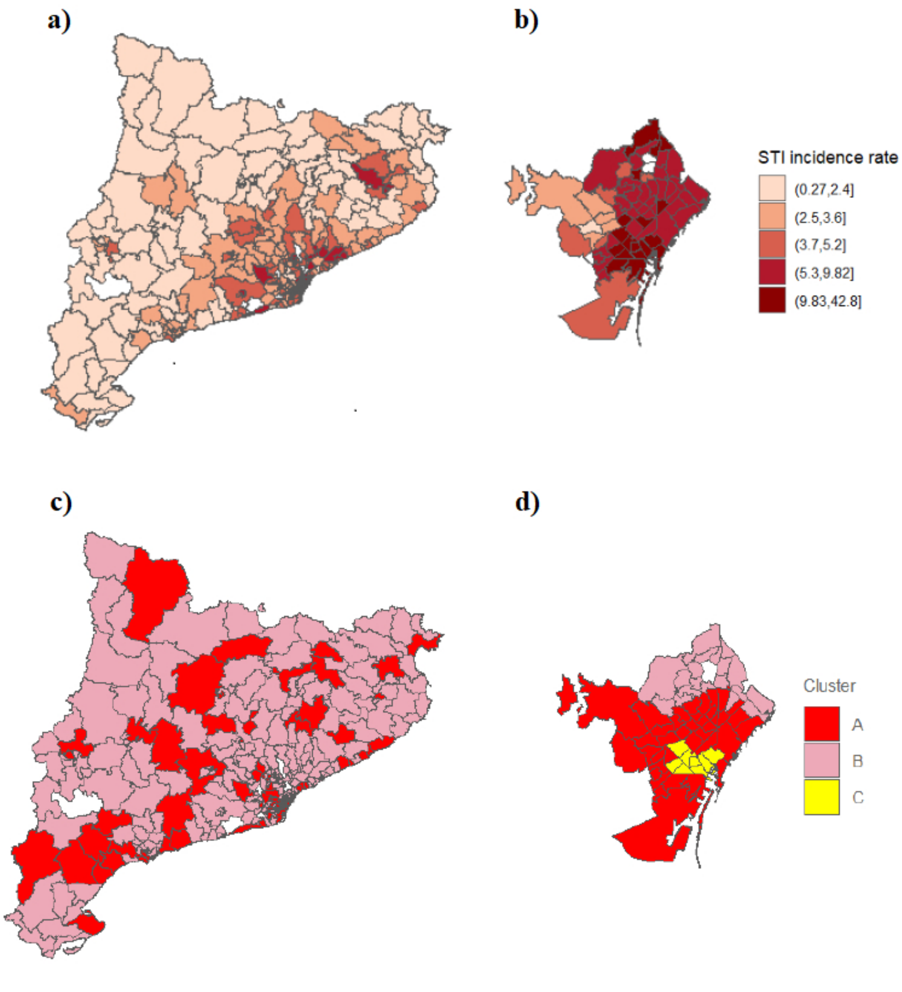
	N TOTAL (N=34,600)	N, HIV + (N=1,376)	N, HIV – (N=33,224)	OR	lower CI	Upper CI	OR adjusted	lower CI	Upper CI
Sex									
Female	14,938	29	14,909	1			1		
Male	19,662	1,347	18,315	37.81	26.69	55.93	23.69	16.67	35.13
Age group									
<20	3,696	5	3,691	1.00			1		
20-29	14,826	328	14,498	16.70	7.70	46.83	8.33	3.82	23.4
30-39	8,704	595	8,109	54.17	25.05	151.57	18.58	8.56	52.13
40-49	4,759	339	4,420	56.62	26.09	158.78	17.66	8.1	49.65
50-59	1,748	89	1,659	39.60	17.80	112.58	13.06	5.84	37.24
>60	867	20	847	17.43	7.04	52.50	6.98	2.8	21.09
Deprivation index[a]									
1st quintile	7,679	501	7,178	1.00			1		
2nd quintile	6,098	210	5,888	0.51	0.43	0.60	0.7	0.59	0.83
3rd quintile	4,163	109	4,054	0.38	0.31	0.47	0.63	0.5	0.78
4th quintile	4,663	186	4,477	0.60	0.50	0.71	0.83	0.69	1
5th quintile	6,347	175	6,172	0.41	0.34	0.48	0.6	0.5	0.72
Missing	5,650	195	5,455	0.51	0.43	0.60	0.51	0.39	0.67
Number of STI (total)									
only episode	29,104	791	28,313	1			1		
2 to 4	5,304	529	4,775	3.96	3.54	4.44	2.69	2.39	3.03
5 to 7	192	56	136	14.74	10.64	20.16	5.96	4.26	8.24
ABS Urbanicity									
Rural	3,699	81	3,618	1			1		
Urban	23,812	1,023	22,789	2	1.61	2.54	1.32	1.04	1.69

Deprivation index[a]: 1st quintile (31.52%), 2nd quintile (40.09%), 3rd quintile (46.27%), 4th quintile (53.98%), 5th quintile (100%)

Table 3. Epidemiological characteristics of the STI socio-epidemiological clusters identified by K-means algorithm, Catalonia, 2017-19.

		Cluster A		Cluster B		Cluster C		Total	
		N	%	N	%	N	%	N	%
Total	Number of episodes	11,527	32.17	19,945	55.66	4,359	12.17	35,831	100.00
	Number of ABS	109	29.62	251	68.21	8	2.17	368	100.00
Number of ABS by quintile of STI incidence rate[a]	1st quintile	710	6.16	1820	9.13	0	0	2530	7.06
	2nd quintile	2136	18.53	3359	16.84	0	0	5495	15.34
	3rd quintile	1377	11.95	6588	33.03	0	0	7965	22.23
	4th quintile	5688	49.35	7508	37.64	0	0	13196	36.83
	5th quintile	1616	14.02	670	3.36	4359	100	6645	18.55
Sex	Men	7,769	67.40	9,379	47.02	3,673	84.26	20,821	58.11
	Women	3,758	32.60	10,566	52.98	686	15.74	15,010	41.89
Age	Total (Med[b] [IQR[c] 95%])	31	18-60	26	17-58	34	20-58	29	17-59
	Men (Med[b] [IQR[c] 95%])	34	19-61	30	18-62	36	21-58	33	19-61
	Women (Med[b] [IQR[c] 95%])	26	17-57	24	16-54	28	18-54.88	24	16-54
Country of birth	Spain	3,920	34.01	6,910	34.65	1,325	30.40	12,155	33.92
	Outside Spain	1,279	11.10	2,819	14.13	435	9.98	4,533	12.65
	Missing	6,328	54.90	10,216	51.22	2,599	59.62	19,143	53.43
Sexual preference	MSM[d]	1,234	10.71	1,104	5.54	655	15.03	2,993	8.35
	MSW	751	6.52	2,164	10.85	64	1.47	2,979	8.31
	WSM	1,440	12.49	6,066	30.41	130	2.98	7,636	21.31
	WSW[e]	75	0.65	295	1.48	6	0.14	376	1.05
	Missing men	5,784	50.18	6,111	30.64	2,954	67.77	14,849	41.44
	Missing women	2,243	19.46	4,205	21.08	550	12.62	6,998	19.53
STI	Gonorrhoea	3,448	29.91	5,240	26.27	1,448	33.22	10,136	28.29
	Chlamydia	5,739	49.79	12,314	61.74	1,649	37.83	19,702	54.99
	Syphilis	2,117	18.37	2,263	11.35	1,027	23.56	5,407	15.09
	LGV	223	1.93	128	0.64	235	5.39	586	1.64
Multiple STI episodes	No	9,927	86.12	18,421	92.36	3,311	75.96	31,659	88.36
	Yes	1,600	13.88	1,524	7.64	1,048	24.04	4,172	11.64
HIV status	HIV negative	10,516	91.23	19,450	97.52	3,673	84.26	33,639	93.88
	HIV positive	1,011	8.77	495	2.48	686	15.74	2,192	6.12
Deprivation index	Med [RIQ 95%]	31.90	3-58.24	44.87	19.21-76.9	25.62	10.68-63.6	39.82	10.68-72.28
ABS urbanicity	Rural	797	6.91	3,158	15.83	0	0.00	3,955	11.04
	Urban	9,461	82.08	15,787	79.15	4,359	100.00	29,607	82.63
	Missing	1,269	11.01	1,000	5.01	0	0.00	2,269	6.33

Quintile of STI incidence rate[a]: 1st quintile (2.41/1.000), 2nd quintile (3.59/1,000), 3rd quintile (5.18/1000), 4th quintile (9.82/1,000), 5th quintile (42.8/1,000), Med[b]: median, RIQ[c]:interquartile range, MSM[d]: men who have sex with men, bisexual men, transgender men, WSW[e]: women who have sex with women, bisexual women, transgender women.



218x238mm (96 x 96 DPI)

Supplementary material

Table S1. HIV/STI epidemiological surveillance system

As instituted by Law 203/2015 (September 15, 2015), the personnel from the Epidemiological Surveillance Network of Catalonia (XVEC) manage the mandatory declaration of diseases and epidemic outbreaks. Jointly with the notification, the health professionals enclose a questionnaire, which includes epidemiological, behavioral, clinical, and geographical variables. The notification comes from two main lines. The first is the Mandatory Declaration of Disease (MDO) system, where a healthcare professional notifies a suspected or confirmed case using the stablished case definition. The notification procedure is done electronically and alternatively by means of the individualized notice form on paper. In compliance with article 13 of law 67/2010 (25 May 2010) of the Health Department of Government of Catalonia, nominal notification of syphilis, gonorrhoea, and LGV have been reported to the MDO since 2006, chlamydia since 2015 and congenital syphilis since 1997. The second line of notification is the Microbiological Notification System of Catalonia (SNMC), which collects microbiological information on selected diseases. Notifications on chlamydia and gonorrhoea are also reported through the SNMC. Notification of new HIV infections was done voluntary between 2001 and 2009, and mandatory and nominal since 2010.

Table S2. Details on study variables.

The socio-demographic variables used in our analysis were sex, age at notification, educational level, deprivation index and country of birth. We used a basic health area (ABS) deprivation index calculated by the Agency for Health Quality and Assessment of Catalonia (AQUAS), attributed to each patient according to their address of residence (categorized in quintiles, first quintile for the ABS with lower deprivation index)[a]. We extracted the classification of ABS as urban or rural (ABS urbanicity) from another deprivation index at ABS level provided by the Primary Health Care Information Systems (SISAP), MEDEA index[b]. Country of birth were categorized by regions adapting for the study those used by WHO. We categorized sexual preference separately for men and women as follows: Men (two groups): MSM (include men who have sex with men, bisexual men and transgender men) and men who have sex with women only (MSW); Women (two groups): WSW (includes women who have sex with women, bisexual women, transgender men) and women who have sex only with men (WSM). Some variables from the epidemiological questionnaire showed high percentages of missing values such are education level, country of birth, and sexual preference (76%, 57%, and 64% respectively, see table 1 and S4). The clinical variables were reinfections, multiple STI episodes (when same persona had more than one during the study period), and coinfection with HIV. Reinfection was defined as more than one episode of the same specific STI during all the study follow-up, but defined differently for each STI, depending on the number of days between successive episodes after the first infection in the same individual; more than 364 for syphilis (although definitive criteria for cure or failure have not been well established yet) and 119 days for gonorrhoea, chlamydia and LGV, respectively[c]. As a geographical variable, we categorized people based on the seven Catalan health regions of their ABS of residence: Alt Pirineu and Aran, Barcelona, Camp de Tarragona, Catalunya central, Girona, Lleida and Terres de l'ebre.

[a] 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.http://observatorisalut.gencat.cat/ca/observatori-desigualtats-salut/indicador_socioeconomic_2015/ (accessed 6 Aug 2020).

[b] Domínguez-Berjón MF, Borrell C, Cano-Serral G, et al. Construcción de un índice de privación a partir de datos censales en grandes ciudades españolas (Proyecto MEDEA). Gac Sanit 2008;22:179–87. doi:10.1157/13123961

[c] CDC - STD Treatment. https://www.cdc.gov/std/treatment/default.htm (accessed 14 Feb 2021).

Table S3. Total number of cases and incidence rates of STI (per 100,000 inhabitants) in Catalonia and its health regions declared at the Catalan epidemiological repository (REC), 2017-19 (N= 42283).

	STI by year					
	2017		2018		2019	
	N	Rate	N	Rate	N	Rate
All Catalonia						
All STI	9,687	129.36	13,724	182.53	18,872	249.29
Chlamydia	3,562	47.57	7,240	96.29	10,400	137.38
Gonorrhoea	3,492	46.63	4,088	54.37	5,782	76.38
Syphilis	2,430	32.45	2,175	28.93	2,370	31.31
LGV	203	2.71	221	2.94	320	4.23
By Health region of residence						
Alt pirineu i Aran	10	14.96	13	19.44	33	49.39
Barcelona	8,205	165.98	11,475	231.16	15,535	310.94
Camp de Tarragona	294	49.69	491	82.54	870	144.85
Catalunya Central	358	69.92	484	94.05	583	112.09
Girona	564	67.29	893	105.90	1,327	155.99
Lleida	212	59.21	247	68.98	404	112.36
Terres de l'Ebre	44	24.70	121	68.34	120	67.99
ABS urbanicity						
Rural	827	11.04	1,375	18.29	1,991	26.30
Urban	6,475	86.47	9,868	131.24	13,626	179.99
Missing	2,385	31.85	2,481	33	3,255	43

Table S4. Distribution of epidemiological characteristics in cases of chlamydia, gonorrhoea, syphilis or lymphogranuloma venerum (LGV) in Catalonia, 2017–2019 (N= 42283).

	All STI		**Chlamydia**		**Gonorrhoea**		**Syphilis**		**LGV**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Education										
Primary school or less	1492	3.53	1034	4.88	334	2.5	4	0.54	120	1.72
Secondary education	5168	12.22	3860	18.21	976	7.3	30	4.03	302	4.33
University	3299	7.8	2450	11.56	591	4.42	31	4.17	227	3.25
Missing	32324	76.45	13858	65.36	11461	85.77	679	91.26	6326	90.7
Country/region of birth										
Spain	13273	31.39	7534	35.53	3890	29.11	282	37.9	1567	22.47
Western countries[a]	537	1.27	246	1.16	157	1.17	17	2.28	117	1.68
North Africa	502	1.19	306	1.44	152	1.14	0	0	44	0.63
Sub-Saharan Africa	193	0.46	123	0.58	53	0.4	3	0.4	14	0.2
Latin America and the Caribbean	3281	7.76	2129	10.04	719	5.38	53	7.12	380	5.45
Eastern Europe and Central Asia	334	0.79	218	1.03	71	0.53	2	0.27	43	0.62
Asia (not central)[b]	216	0.51	145	0.68	50	0.37	1	0.13	20	0.29
Missing	23947	56.64	10501	49.53	8270	61.89	386	51.88	4790	68.67
Health region of residence										
Alt pirineu i Aran	56	0.13	30	0.14	14	0.1	0	0	12	0.17
Barcelona	35215	83.28	17108	80.69	11566	86.56	708	95.16	5833	83.63
Camp de Tarragona	1655	3.91	930	4.39	390	2.92	10	1.34	325	4.66
Catalunya central	1425	3.37	861	4.06	371	2.78	8	1.08	185	2.65
Girona	2784	6.58	1595	7.52	730	5.46	8	1.08	451	6.47
Lleida	863	2.04	499	2.35	241	1.8	5	0.67	118	1.69
Terres de l'ebre	285	0.67	179	0.84	50	0.37	5	0.67	51	0.73

Western countries[a]: Western Europe, North America, Australia, and New Zealand, Asia (not central)[b]: South-eastern Asia, Southern Asia, and Western Asia.

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract (pages 4 and 7)
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found (page 4)
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported (page 6 and 7)
Objectives	3	State specific objectives, including any prespecified hypotheses (page 7)
Methods		
Study design	4	Present key elements of study design early in the paper (page 7)
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection (page 7 and 8)
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up (page 7)
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable (page 7-9, supplementary material tables S1 and S2)
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group (page 7-9, supplementary material tables S1 and S2)
Bias	9	Describe any efforts to address potential sources of bias (page 8 and 9 and table S2 in supplementary material)
Study size	10	Explain how the study size was arrived at (page 7 and 8 and table S1 in supplementary material)
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why (page 7-9, supplementary material tables S1 and S2)
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) If applicable, explain how loss to follow-up was addressed
		(*e*) Describe any sensitivity analyses
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (page 8 and 9 and table S3 and S4 in supplementary material)
		(b) Indicate number of participants with missing data for each variable of interest (in all tables)

		(c) Summarise follow-up time (eg, average and total amount) (page 7)
Outcome data	15*	Report numbers of outcome events or summary measures over time (page 10-12 and all tables)
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (page 8 and 9 and table 2, for unadjusted estimates)
		(*b*) Report category boundaries when continuous variables were categorized (in all tables)
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives (page 10-12)
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias (page 15)
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence (page 13-15)
Generalisability	21	Discuss the generalisability (external validity) of the study results (page 12-16)
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based (page 16)

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

BMJ Open

STI epidemic re-emergence, socioepidemiological clusters identification and characterisation, and factors associated with HIV coinfection in Catalonia, Spain, during 2017–2019: a retrospective population-based cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-052817.R1
Article Type:	Original research
Date Submitted by the Author:	16-Aug-2021

Complete List of Authors:	Sentís, Alexis; CEEISCAT; Epiconcept SAS, Epidemiology Montoro-Fernandez, Marcos; CEEISCAT; IGTP Lopez-Corbeto, E; CEEISCAT; CIBERESP Egea-Cortés, Laia; CEEISCAT; IGTP Nomah, DK ; CEEISCAT; Autonomous University of Barcelona, Department of Paediatrics, Obstetrics and Gynecology and Preventive Medicine Díaz, Yesika; CEEISCAT; IGTP Garcia de Olalla, Patricia; Public Health Agency of Barcelona, Epidemiology; CIBERESP Mercuriali, Lilas; Public Health Agency of Barcelona, Epidemiology Borrell, Núria; Public Health Agency of Catalonia, Epidemiological Surveillance and Response to Public Health Emergencies Service in Tarragona Reyes-Urueña, Juliana; CEEISCAT; CIBERESP Casabona, Jordi; ASPC, CEEISCAT; CIBERESP

Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	HIV/AIDS, Public health, Research methods, Infectious diseases
Keywords:	Epidemiology < INFECTIOUS DISEASES, HIV & AIDS < INFECTIOUS DISEASES, SEXUAL MEDICINE, Infection control < INFECTIOUS DISEASES, Public health < INFECTIOUS DISEASES, PREVENTIVE MEDICINE

Title: STI epidemic re-emergence, socioepidemiological clusters identification and characterisation, and factors associated with HIV coinfection in Catalonia, Spain, during 2017–2019: a retrospective population-based cohort study

Authors: Alexis Sentís,[1,2,3] Marcos Montoro-Fernandez,[1,4] Evelin Lopez-Corbeto,[1,4,5] Laia Egea-Cortés,[1,4] Daniel Kwakye Nomah,[1,6] Yesika Díaz,[1,4] Patrícia Garcia de Olalla,[5,7] Lilas Mercuriali,[7] Núria Borrell,[8] Juliana Reyes-Urueña,[1,4,5] Jordi Casabona,[1,4,5,6] and the Catalan HIV and STI surveillance group*

Affiliations:

[1]Department of Health, Centre of Epidemiological Studies of Sexually Transmitted Disease and AIDS in Catalonia (CEEISCAT), Generalitat of Catalonia, Badalona, Spain

[2]Pompeu Fabra University (UPF), Barcelona, Spain

[3] Epidemiology Department, Epiconcept, Paris, France

[4]Fundació Institut d'Investigació Germans Trias i Pujol (IGTP), Badalona, Spain.

[5]Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain.

[6]Department of Paediatrics, Obstetrics and Gynecology and Preventive Medicine, Universitat Autònoma de Barcelona, Badalona, Spain.

[7]Epidemiology Service. Public Health Agency of Barcelona, Barcelona, Spain.

[8]Epidemiological Surveillance and Response to Public Health Emergencies Service in Tarragona, Agency of Public Health of Catalonia, Generalitat of Catalonia, Tarragona, Spain

*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 Mercuriali, 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 Martinez, 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 Mendioroz, 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).

Correspondence to:

Alexis Sentís Fuster

Address: Fundació Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol (IGTP), Edifici Muntanya, Carretera de Can Ruti, Camí de les Escoles s/n, 08916 Badalona (Spain)

E-mail address: alexissentis@gmail.com,

Journal: BMJ Open

- Article type**:** Original article
- Word count: 3992
- Figures and tables: 1 figure, 4 tables (max: 5 tables and/or figures)
- References: 30
- Supplementary material: 5 tables

Key words: Epidemiology, infection control, public health, HIV & AIDS, sexual medicine, preventive medicine

ABSTRACT

Objectives: To describe the epidemiology of sexually transmitted infections (STIs), identify and characterise socioepidemiological 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 (LGV) 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 socioepidemiological 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 2017-2019 (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 odds ratio [aOR] 23.69; 95% confidence interval [CI] 16.67–35.13), age 30–39 years (versus <20 years, aOR 18.58; 95% CI 8.56–52.13), having 5–7 STI episodes (versus 1 episode, aOR 5.96; 95% CI 4.26–8.24) and living in urban areas (aOR 1.32; 95% CI 1.04–1.69). Living in the most deprived BHAs (aOR 0.60; 95% CI 0.50–0.72) 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 MSM in urban and less deprived areas showed higher rates of incidence, multiple STI episodes, and HIV coinfection.

Conclusions: We recommend socioepidemiological identification and characterization of STI clusters and factors associated with HIV coinfection to identify at-risk populations at a health area level to design effective interventions.

Word count: 300 (max: 300 words)

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 characterize the re-emergence of STIs, perform socioepidemiological clustering and reveal factors associated with HIV co-infection.
- To our knowledge, this is the first study to apply the k-means clustering methodology to identify and characterize distinct socioepidemiological clusters of STI at a small health area level.
- MSM, heterosexual women and young adults should be considered priority target populations for preventative strategies of STI and HIV, taking into account structural and social determinants that were identified as crucial in this analysis.
- 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.

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.[1,2]

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.[3–6] 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.[7,8] 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.[9] 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.[7,9] 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 globalization.[10]

STIs and HIV infections are overlapping epidemics, which, apart 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.[11,12] Some studies have described social determinants of health, discrimination and inequalities as the main factors associated with the spatiotemporal clustering of STI cases.[13,14] 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 sociological and epidemiological factors.[15] The socioepidemiological characterization 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 socioepidemiological 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,[16] 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 supplementary table S1). Case definitions for surveillance reporting are standardized according to the European Union definitions established by the European Centre for Disease Prevention and Control (online supplementary table S2).[17,18]

Analysis variables

We extracted data around epidemiological, sociodemographic and clinical variables as detailed in online supplementary 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, Catalan HIV/STI Registry surveillance team performs duplicate checks 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 anonymized 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 5,000–25,000 people. The socioeconomic level of the BHAs were 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 categorized in quintiles, with the first quintile being the least deprived.[19]

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 supplementary Table S3). As information regarding treatment response was not available, episodes occurring outside of the specific timeframes 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 heterogenous units (in our case BHAs) into clusters based on similarities in characteristics (variable values and categorical distribution among the BHA).[15]

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 intra-cluster variation is minimum. Our external validation process was based on a description of the possible socioepidemiological variables to determine the most appropriate clustering algorithm for our dataset. 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 socioepidemiological 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 summarize epidemiological, clinical, sociodemographic and geographical variables for the total confirmed STI cases and by STI clusters. Continuous variables were summarised as median (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 supplementary 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 1,000 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 (version 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 de-identified 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 STIs, 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 9,687 in 2017 to 18,872 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 ≥30 years). Indeed, women under 30 years of age presented the highest decrease in both number of cases, with an increase of 155.8% vs 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) vs 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% 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, 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 supplementary 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 greater among males (aOR 23.69; 95% CI 16.67–35.13 compared with females) and in urban BHAs (aOR 1.32; 95% CI 1.04–1.69 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 socioepidemiological 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. 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 socioepidemiological 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 versus 160.6 per 100,000 population), but represented the majority of all reported STI cases (55.7%). STI-infected individuals in Cluster 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% versus 11.0%) and had a higher proportion of heterosexual men and women (approximately 12% higher) and chlamydia cases (61.7% versus 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 100,000 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 4,359 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 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 socioepidemiological clusters of STI showed that young women living in rural and more deprived areas were more likely to be affected by chlamydia. Further, 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.[3–6] The rise in STI cases has been partially attributed to enhancement of surveillance systems and the introduction of improved diagnostic tools in recent years.[10] 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.[20–22]

Chlamydia has been reported more frequently in WSM, while syphilis, gonorrhoea and LGV were more common in MSW and MSM.[3–9] 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 of STIs in Catalonia in 2007–2015,[12] 2012-2017[23] and 2018-2019,[9] 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.[24]

Consistent with previous data,[12] 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.[25] The World Health Organization and other public health agencies have emphasized 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 characterization of populations at increased risk of STIs.[10,25] Sociodemographic and socioeconomic are increasingly being established as more important risk factors of STI acquisition than individual behaviours, particularly among women from disadvantaged groups. [26,27]

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 heterogenous variables such as sociodemographic, clinical, cognitive, vital signs and laboratory parameters.[28] More recently, its potential to monitor and group trends of SARS-CoV-2 prevalence by magnitude (higher, medium and lower) at a regional level in Italy has been described.[29] Identification of these 'clusters of characteristics' may be useful, in their specific context, to better detect and characterize 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.[30] To the best of our knowledge, this current study is the first to apply the k-means clustering methodology to identify and characterise socioepidemiological 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.[3–6,9,12,24] The age category above 60 years old may contribute to residual cofounding 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 characterization 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 not only an issue of the health sector, 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, not only among the MSM population, but 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 prevalence as the key monitoring variable.

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.

Competing interests

All authors declare no potential conflicts of interest relevant to this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Contributors

AS conceptualized and designed the study. MMF cleaned the database, MMF, LEC and YD performed the statistical and cluster analysis. AS, ELC and DN reviewed scientific literature, AS, JRU, and JC drafted the manuscript and AS, ELC, PGO, LM, NB, JRU and JC interpreted the results. All authors critically reviewed the manuscript and approved the final version to be published.

Data sharing statement

No additional data are available.

REFERENCES

1. 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;97:548-562P.

2. WHO Department of Reproductive Health and Research. Report on global sexually transmitted infection surveillance, 2018. Available at: https://apps.who.int/iris/bitstream/handle/10665/277258/9789241565691-eng.pdf?ua=1 [Accessed 8 January 8 2021].

3. European Centre for Disease Prevention and Control. Chlamydia infection. In: ECDC. Annual epidemiological report for 2018. Stockholm: ECDC; 2020. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/AER-for-2018-STI-chlamydia.pdf [Accessed 7 August 2021].

4. European Centre for Disease Prevention and Control. Gonorrhoea. In: ECDC. Annual epidemiological report for 2018. Stockholm: ECDC; 2020. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/gonorrhoea-annual-epidemiological-report-2018.pdf [Accessed 7 August 2021].

5. European Centre for Disease Prevention and Control. Syphilis. In: ECDC. Annual epidemiological report for 2018. Stockholm: ECDC; 2020. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/syphilis-aer-2018.pdf [Accessed 7 August 2021].

6. European Centre for Disease Prevention and Control. Lymphogranuloma venereum. In: ECDC. Annual epidemiological report for 2018. Stockholm: ECDC; 2020. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/lymphogranuloma-venereum-aer-2018.pdf [Accessed 7 August 2021].

7. Centro Nacional de Epidemiología. Unidad de vigilancia del VIH y conductas de riesgo. Vigilancia epidemiológica 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, Calidad e Innovación; 2019. Available at: https://www.mscbs.gob.es/ciudadanos/enfLesiones/enfTransmisibles/sida/vigilancia/Vigilancia_ITS_1995_2017_def.pdf [Accessed 27 March 2021]

8. Área de vigilancia del VIH y conductas de riesgo. Vigilancia epidemiológica de las infecciones de transmisión sexual, 1995-2010. Madrid: Centro Nacional de Epidemiología; 2012. Available at:

https://www.mscbs.gob.es/ciudadanos/enfLesiones/enfTransmisibles/sida/vigilancia/VigilanciaITS1995_2010.pdf [Accessed 27 March 2021].

9. Centre d'Estudis Epidemiològics sobres 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: CEEISACT; 2020. Available at: https://canalsalut.gencat.cat/web/.content/_A-Z/S/sida/enllasos/anual_ITS.pdf [Accessed 27 March 2021].

10. Taylor MM, Wi TEC. Transforming and integrating STI surveillance to enhance global advocacy and investment in STI control. *J Int AIDS Soc* 2019;22:25361.

11. Centers for Disease Control and Prevention. HIV/AIDS & STDs.Available at: https://www.cdc.gov/std/hiv/default.htm. [Accessed 22 February 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; 2015. Available at: https://scientiasalut.gencat.cat/bitstream/handle/11351/3418/informe_SIVES_2015_informe_epidemiologic_CEEISCAT_2015.pdf.pdf?sequence=1&isAllowed=y [Accessed 27 March 2021].

17. Notifiable diseases and epidemic outbreaks. Department of Health, Public Health Agency of Catalonia, Generalitat de Catalunya; 2016. Available at: https://canalsalut.gencat.cat/ca/professionals/vigilancia-epidemiologica/malalties-de-declaracio-obligatoria-i-brots-epidemics/ [Accessed 20 June 2021].

18. European Centre for Disease Prevention and Control. EU case definitions; 2018. Available at: https://www.ecdc.europa.eu/en/surveillance-and-disease-data/eu-case-definitions. [Accessed 7 August 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 at: http://observatorisalut.gencat.cat/ca/observatori-desigualtats-salut/indicador_socioeconomic_2015/ [Accessed 6 August 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. 2019;18:1–11.

22. Jansen K, Steffen G, Potthoff A, *et al.* STI in times of PrEP: High prevalence of chlamydia, gonorrhea, and mycoplasma at different anatomic sites in men who have sex with men in Germany. *BMC Infect Dis* 2020;20:110.

23. Vives N, Olalla PG de, 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, Gonorrhea 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. *Canada Commun Dis Rep* 2019;45:63–6.

26. Hill A V., 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 Heal* 2019;64:319–26.

27. Jagadesh S, Combe M, Couppié P *et al.* Mapping priority neighborhoods: A novel approach to cluster identication in HIV/AIDS population. *Res Sq* 2020 [doi: 10.21203/rs.3.rs-17843/v1].

28. de la Fuente-Tomas L, Arranz B, Safont G, *et al.* Classification of patients with bipolar disorder using k-means clustering. *PLoS One* 2019;14.

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;17:1–14.

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.

FIGURE LEGEND

Figure 1. Incidence rates (per 1,000 population) and socioepidemiological clusters of STIs by BHA during 2017–2019. a) STI incidence rates in Catalonia; b) STI incidence rates in Barcelona city*; c) STI socioepidemiological clusters in Catalonia, and d) STI socioepidemiological 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 visualization of Cluster C. From a total of 373 Catalan BHA five (Garraf rural, Polinyà-Sentmenat, Ribes-Olivella. Roquetes-Canyelles, Viladecans 3) were excluded from the K-means clustering analysis because their delimitations and populations changed during the study period.

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

Characteristic	All STIs	Chlamydia	Gonorrhoea	Syphilis	LGV
Total cases, n (%)	42,283 (100)	21,202 (50.1)	13,362 (31.6)	6,975 (16.5)	744 (1.8)
Sex, n (%)					
Female	16,676 (39.4)	13,125 (61.9)	2,667 (20.0)	875 (12.5)	9 (1.2)
Male	25,607 (60.6)	8,077 (38.1)	10,695 (80.0)	6,100 (87.5)	735 (98.8)
Age, n (%)					
<20 years	4,438 (10.5)	3,311 (15.6)	984 (7.4)	137 (2.0)	6 (0.8)
20–29 years	17,691 (41.8)	10,707 (50.5)	5,361 (40.1)	1,462 (21.0)	161 (21.6)
30–39 years	11,102 (26.3)	4,454 (21.0)	4,116 (30.8)	2,242 (32.1)	290 (39.0)
40–49 years	6,092 (14.4)	2,087 (9.8)	2,032 (15.2)	1,757 (25.2)	216 (29.0)
50–59 years	2,037 (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)
Sexual preference, n (%)					
MSM*	3,270 (7.7)	785 (3.7)	1,321 (9.9)	993 (14.2)	171 (23.0)
MSW	3,149 (7.5)	1,863 (8.8)	1,040 (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	8,189 (19.4)	7,034 (33.2)	966 (7.2)	186 (2.7)	3 (0.4)
Missing (male)	19,188 (45.4)	5,429 (25.6)	8,334 (62.4)	4,864 (69.7)	561 (75.4)
Missing (female)	8,072 (19.1)	5,756 (27.2)	1,632 (12.2)	679 (9.7)	5 (0.7)
STI reinfection, n (%)	4,558 (10.8)	1,418 (6.7)	2,098 (15.7)	955 (13.7)	87 (11.7)
HIV coinfection, n (%)	2,443 (5.8)	467 (2.2)	897 (6.7)	893 (12.8)	186 (25.0)
Deprivation index‡					
First quintile (least deprived)	10,271 (24.3)	5,185 (24.5)	3,040 (22.8)	1,757 (25.2)	289 (38.8)
Second quintile	7,465 (17.7)	4,328 (20.4)	2,012 (15.1)	1,037 (14.9)	88 (11.8)
Third quintile	4,859 (11.5)	2,763 (13.0)	1,332 (10.0)	716 (10.3)	48 (6.5)
Fourth quintile	5,703 (13.5)	3,217 (15.2)	1,578 (11.8)	827 (11.9)	81 (10.9)
Fifth quintile	7,689 (18.2)	4,319 (20.4)	2,211 (16.6)	1,079 (15.5)	80 (10.8)
Missing	6,296 (14.9)	1,390 (6.6)	3,189 (23.9)	1,559 (22.4)	158 (21.2)
Health region of residence, n (%)					
Barcelona	35,215 (83.3)	17,108 (80.7)	11,566 (86.6)	5,833 (83.6)	708 (95.2)
Other regions	7,068 (16.7)	4,094 (19.3)	1,796 (13.4)	1,142 (16.4)	36 (4.8)
BHA setting					
Rural	4,193 (9.9)	2,614 (12.3)	1,039 (7.8)	516 (7.4)	24 (3.2)
Urban	29,969 (70.9)	16,347 (77.1)	8,566 (64.1)	4,516 (64.8)	540 (72.6)
Missing	8,121 (19.2)	2,241) (10.6)	3,757) (28.1)	1,943) (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.

‡1st quintile (31.52%), 2nd quintile (40.09%), 3rd quintile (46.27%), 4th quintile (53.98%), 5th 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.

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	
Total							
Total	9,687	128.2	13,724	180.6	18,872	245.9	<0.001
Sex							
Female[a]	3,362	87.4	5,503	142.2	7,811	200.0	<0.001
Male[b]	6,325	170.5	8,221	220.4	11,061	293.4	<0.001
Age, years							
<30[c]	4,607	198.4	7,181	306.9	10,341	436.5	<0.001
≥30[d]	5,080	97.1	6,543	124.4	8,531	160.8	<0.001
Sex and age							
Male <30 years[f]	2,411	202.9	3,347	279.5	4,668	384.0	<0.001
Male ≥30 years[g]	3,914	155.2	4,874	192.4	6,393	250.3	<0.001
Female <30 years[h]	2,196	193.6	3,834	335.7	5,673	491.9	<0.001
Female ≥30 years[I]	1,166	43.0	1,669	61.2	2,138	77.7	<0.001
STI type (total Catalonia[j])							
Chlamydia	3,562	47.1	7,240	95.3	10,400	135.5	<0.001
Gonorrhoea	3,492	46.2	4,088	53.8	5,782	75.3	<0.001
Syphilis	2,430	32.2	2,175	28.6	2,370	30.9	0.1572
LGV	203	2.7	221	2.9	320	4.2	<0.001
Health region of residence							
Alt pirineu i Aran[k]	10	13.9	13	18.1	33	45.7	<0.001
Barcelona[L]	8,205	165.0	11,475	229.5	15,535	307.8	<0.001
Camp de Tarragona[m]	294	49.1	491	81.3	870	142.2	<0.001
Catalunya Central[n]	358	69.4	484	93.1	583	110.7	<0.001
Girona[o]	564	65.7	893	103.2	1,327	151.5	<0.001
Lleida[p]	212	58.9	247	68.5	404	111.5	<0.001
Terres de l'Ebre[q]	44	24.5	121	67.7	120	67.2	<0.001
BHA setting[a]							
Rural	827	NA	1,375	NA	1,991	NA	NA
Urban	6,475	NA	9,868	NA	13,626	NA	NA
Missing	2,385	NA	2,481	NA	3,255	NA	NA

BHA, Basic Health Area; LGV, lymphogranuloma venereum; STI, sexually transmitted infection. Data source for denominators (a-q): the Statistical Institute of Catalonia (IDESCAT): online supplementary Table S4.

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=1,376)	OR	95% CI	aOR	95% CI
Sex						
Female	14,938	29	1 (ref)		1 (ref)	
Male	19,662	1,347	37.81	26.69–55.93	23.69	16.67–35.13
Age group, years						
<20	3,696	5	1 (ref)		1 (ref)	
20-29	14,826	328	16.70	7.70–46.83	8.33	3.82–23.40
30-39	8,704	595	54.17	25.05–151.57	18.58	8.56–52.13
40-49	4,759	339	56.62	26.09–158.78	17.66	8.10–49.65
50-59	1,748	89	39.60	17.80–112.58	13.06	5.84–37.24
>60	867	20	17.43	7.04–52.50	6.98	2.80–21.09
Deprivation index*						
First quintile (least deprived)	7,679	501	1 (ref)		1 (ref)	
Second quintile	6,098	210	0.51	0.43–0.60	0.70	0.59–0.83
Third quintile	4,163	109	0.38	0.31–0.47	0.63	0.50–0.78
Fourth quintile	4,663	186	0.60	0.50–0.71	0.83	0.69–1.00
Fifth quintile	6,347	175	0.41	0.34–0.48	0.60	0.50–0.72
Missing	5,650	195	0.51	0.43–0.60	0.51	0.39–0.67
STI episodes (total), n						
1	29,104	791	1 (ref)		1 (ref)	
2–4	5,304	529	3.96	3.54–4.44	2.69	2.39–3.03
5–7	192	56	14.74	10.64–20.16	5.96	4.26–8.24
BHA setting						
Rural	3,699	81	1 (ref)		1 (ref)	
Urban	23,812	1,023	2	1.61–2.54	1.32	1.04–1.69

*1st quintile (31.52%), 2nd quintile (40.09%), 3rd quintile (46.27%), 4th quintile (53.98%), 5th quintile (100%).

aOR, adjusted odds ratio; BHA, Basic Health Area; CI, confidence interval; LGV, lymphogranuloma venereum; MSM, men who have sex with men; MSW, men who have sex with women; OR, odds ratio; STI, sexually transmitted infection; WSM, women who have sex with men; WSW, women who have sex with women.

Table 4. Characteristics of socioepidemiological STI clusters in Catalonia, Spain (2017–2019)

Characteristic	Cluster A	Cluster B	Cluster C	Total*
Demographics				
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
Reported STI cases, n (%)				
Total	11,527 (32.2)	19,945 (55.7)	4,359 (12.2)	35,831 (100)
Sex				
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)
Country of birth				
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)
Sexual preference				
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)
STI type				
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)
STI incidence rate categories of BHAs				
First quintile (2.4 per 1,000)	710 (6.2)	1820 (9.1)	0	2530 (7.1)
Second quintile (3.6 per 1,000)	2136 (18.5)	3359 (16.8)	0	5495 (15.3)
Third quintile (5.2 per 1,000)	1377 (12.0)	6588 (33.0)	0	7965 (22.2)
Fourth quintile (9.8 per 1,000)	5688 (49.4)	7508 (37.6)	0	13196 (36.8)
Fifth quintile (42.8 per 1,000)	1616 (14.0)	670 (3.4)	4359 (100)	6645 (18.6)

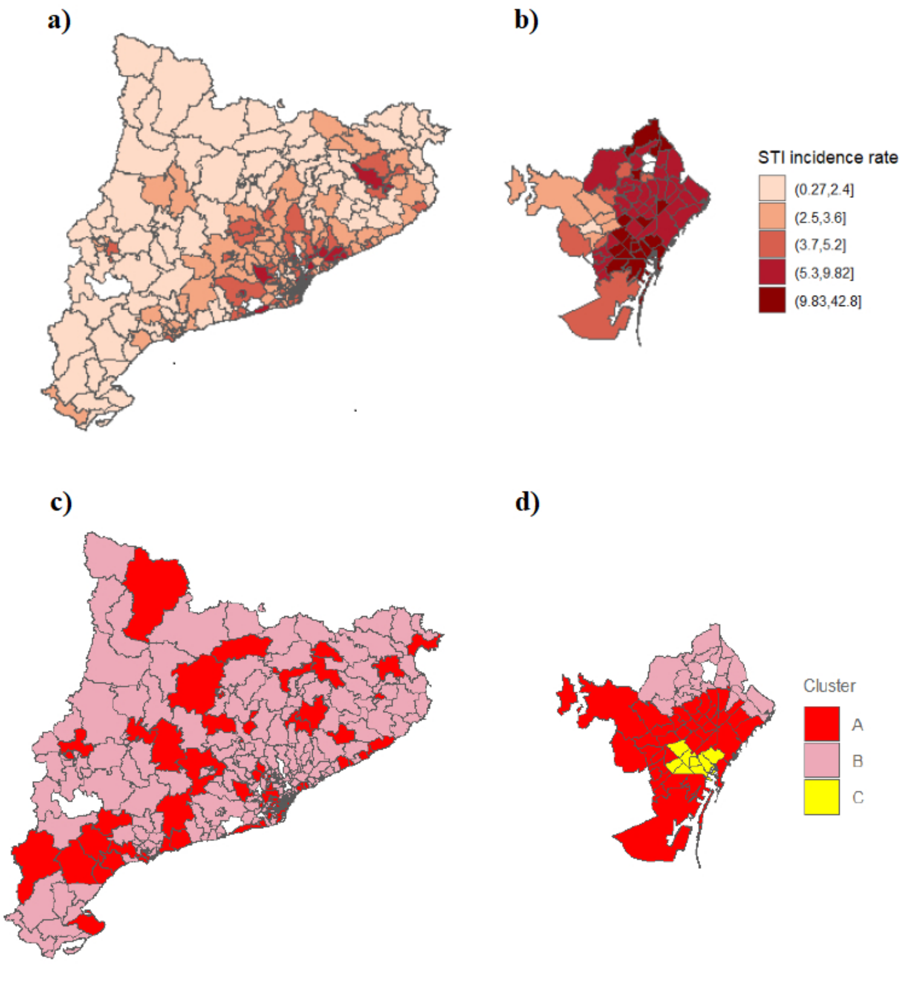
BHA setting				
Rural	797 (6.9)	3,158 (15.8)	0	3,955 (11.0)
Urban	9,461 (82.1)	15,787 (79.2)	4,359 (100)	29,607 (82.6)
Missing	1,269 (11.0)	1,000 (5.0)	0	2,269 (6.3)

*Of the 373 Catalan BHAs, five (Garraf rural, Polinyà-Sentmenat, Ribes-Olivella, Roquetes-Canyelles, 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; 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.



218x238mm (96 x 96 DPI)

SUPPLEMENTARY MATERIAL

Table S1. Epidemiological Repository of Catalonia (REC, in Catalan), an electronic registry used by the Epidemiological Surveillance Network of Catalonia (XVEC, in Catalan)

As instituted by Law 203/2015 (September 15, 2015), personnel from the Epidemiological Surveillance Network of Catalonia (XVEC) manage the mandatory declaration of diseases and epidemic outbreaks. Along with the notification, healthcare professionals enclose a questionnaire that describes epidemiological, behavioural, clinical, and geographical parameters. The notification comes from two main sources. The first is the Mandatory Declaration of Disease (MDO) system, where a healthcare professional reports a suspected or confirmed case using established case definitions. The notification procedure is done electronically or, alternatively, by means of the individualized notice form on paper. In compliance with article 13 of law 67/2010 (25 May 2010) of the Health Department of Government of Catalonia, nominal notification of syphilis, gonorrhoea, and LGV have been reported to the MDO since 2006, chlamydia since 2015 and congenital syphilis since 1997. The second source of notification is the Microbiological Notification System of Catalonia (SNMC), which collects microbiological information on selected diseases. Notifications on chlamydia and gonorrhoea are also reported through the SNMC. Notification of new HIV infections was done on a voluntary basis between 2001 and 2009, and mandatory and nominal since 2010.

Table S2. STI case definitions from the Public Health Agency of Catalonia, Department of Health, Generalitat de Catalunya.

CHLAMYDIA: **1) Laboratory criteria for diagnosis:** Isolation of *Chlamydia trachomatis* by culture in a sample of the genitourinary tract, anal or conjunctiva, or clinical sample; or demonstration of *C. trachomatis* by detection of specific antigens or by direct immunofluorescence (DFA) in a clinical sample; or detection of specific genomic fragments of *C. trachomatis* in a clinical specimen. **2) Confirmed case:** Person with compatible laboratory criteria. **3) Probable case:** Person with clinically compatible criteria, especially if it is epidemiologically related.

GONORRHOEA: **1) Laboratory criteria for diagnosis:** Isolation by culture of *Neisseria gonorrhoeae* in a clinical specimen, or detection of specific genomic fragments of *N. gonorrhoeae* in a clinical specimen, or microscopic detection of gram-negative intracellular diplococci in urethral exudates in men. **2) Confirmed case:** Person with compatible laboratory criteria. **3) Probable case:** Person with clinically compatible criteria, especially if it is epidemiologically related.

SYPHILIS: **1) Laboratory criteria for diagnosis**: Demonstration of *Treponema pallidum* by dark field microscopy, by direct immunofluorescence (DFA), of genomic fragments, in lesion secretions. Detection of antibodies against *T. pallidum* by specific tests (TPHA, TPPA or EIA) and, in addition, one of the following methods: FTA-ABS, EIA immunotransference, non-specific reactive serological test (VDRL, RPR), detection of IgM antibodies -TP. **2) Confirmed case**: Person with compatible laboratory criteria. **3) Probable case:** Person with clinically compatible criteria, especially if it is epidemiologically related.

LGV: **1) Laboratory criteria for diagnosis:** Detection of specific genomic fragments of *C. trachomatis* in a clinical sample, and in addition Identification of serovar L1, L2 or L3. **2) Confirmed case:** Person with compatible laboratory criteria. **3) Probable case:** Person with clinically compatible criteria, especially if it is epidemiologically related.

Notifiable diseases and epidemic outbreaks. Department of Health, Public Health Agency of Catalonia, Generalitat de Catalunya. Available at: https://canalsalut.gencat.cat/ca/professionals/vigilancia-epidemiologica/malalties-de-declaracio-obligatoria-i-brots-epidemics/ ((accessed 20 Jun 2021).

Table S3. Details on study variables.

The socio-demographic variables used in our analysis were sex, age at notification, educational level, deprivation index and country of birth. We used a basic health area (ABS) deprivation index calculated by the Agency for Health Quality and Assessment of Catalonia (AQUAS), attributed to each patient according to their address of residence (categorized in quintiles, first quintile for the ABS with lower deprivation index)[a]. We extracted the classification of ABS as urban or rural (ABS urbanicity) from another deprivation index at ABS level provided by the Primary Health Care Information Systems (SISAP), MEDEA index[b]. Country of birth were categorized by regions adapting for the study those used by WHO. We categorized sexual preference separately for men and women as follows: Men (two groups): MSM (include men who have sex with men, bisexual men and transgender men) and men who have sex with women only (MSW); Women (two groups): WSW (includes women who have sex with women, bisexual women, transgender men) and women who have sex only with men (WSM). Some variables from the epidemiological questionnaire showed high percentages of missing values such are education level, country of birth, and sexual preference (76%, 57%, and 64% respectively, see table 1 and S4). The clinical variables were reinfections, multiple STI episodes (when same persona had more than one during the study period), and coinfection with HIV. Reinfection was defined as more than one episode of the same specific STI during all the study follow-up, but defined differently for each STI, depending on the number of days between successive episodes after the first infection in the same individual; more than 364 for syphilis (although definitive criteria for cure or failure have not been well established yet) and 119 days for gonorrhoea, chlamydia and LGV, respectively[c]. As a geographical variable, we categorized people based on the seven Catalan health regions of their ABS of residence: Alt Pirineu and Aran, Barcelona, Camp de Tarragona, Catalunya central, Girona, Lleida and Terres de l'ebre.

[a] 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.http://observatorisalut.gencat.cat/ca/observatori-desigualtats-salut/indicador_socioeconomic_2015/ (accessed 6 Aug 2020).

[b] Domínguez-Berjón MF, Borrell C, Cano-Serral G, et al. Construcción de un índice de privación a partir de datos censales en grandes ciudades españolas (Proyecto MEDEA). Gac Sanit 2008;22:179–87. doi:10.1157/13123961

[c] CDC - STD Treatment. https://www.cdc.gov/std/treatment/default.htm (accessed 14 Feb 2021).

Table S4. Denominators for STI incidence rates calculations.

Data source for denominators (a-q) in Table 2 (main text): the Statistical Institute of Catalonia (IDESCAT) [data provided by IDESCAT on 23 June 2020]:

[a]Total female population in Catalonia: 3,845,630 in 2017, 3,869,739 in 2018, and 3,905,094 in 2019.

[b]Total male population in Catalonia: 3,710,200 in 2017, 3,730,326 in 2018, and 3,770,123 in 2019.

[c]Total population <30 yrs in Catalonia: 2,322,227 in 2017, 2,339,673 in 2018, and 2,368,830 in 2019.

[d]Total population ≥ 30 yrs in Catalonia: 5,233,603 in 2017, 5,260,392 in 2018, and 5,306,387 in 2019.

[e]Total male population<30 yrs in Catalonia: 1,187,850 in 2017, 1,197,499 in 2018, and 1,215,583 in 2019.

[f]Total male population≥ 30 yrs in Catalonia: 2,522,350 in 2017, 2,532,827 in 2018, and 2,554,540 in 2019.

[g]Total female population<30yrs in Catalonia: 1,134,377 in 2017, 1,142,174 in 2018, and 1,153,247 in 2019.

[h]Total female population≥ 30yrs in Catalonia: 2,711,253 in 2017, 2,727,565 in 2018, and 2,751,847 in 2019.

[I]Total population in Catalonia: 7,555,830 in 2017, 7,600,065 in 2018, and 7,675,217 in 2019. Used as a denominator to compute STI rates, total and by disease, and rates according ABS urbanicity category.

[j]Total population in Alt pirineu i Aran: 71,958 in 2017, 71,888 in 2018, and 72,276 in 2019.

[k]Total population in Barcelona: 4,972,179 in 2017, 5,000,125 in 2018, and 5,047,597 in 2019.

[L]Total population in Camp de Tarragona: 598,683 in 2017, 603,743 in 2018, and 611,950 in 2019.

[m]Total population in Catalunya Central: 515,578 in 2017, 519,819 in 2018, and 526,544 in 2019.

[n]Total population in Girona: 857,877 in 2017, 865,282 in 2018, and 875,722 in 2019.

[o]Total population in Lleida: 359,729 in 2017, 360,497 in 2018, and 362,428 in 2019.

[p]Total population in Terres de l'Ebre: 179,826 in 2017, 178,711 in 2018, and 178,700 in 2019.

[q]NA=Not available denominators.

Table S5. Distribution of epidemiological characteristics in cases of chlamydia, gonorrhoea, syphilis or lymphogranuloma venerum (LGV) in Catalonia, 2017–2019 (N= 42,283).

	All STI (N=42,283)		Chlamydia (N=21,202)		Gonorrhoea (N=13,362)		Syphilis (N=6,975)		LGV (N =744)	
	N	%	N	%	N	%	N	%	N	%
Education										
Primary school or less	1492	3.53	1034	4.88	334	2.5	4	0.54	120	1.72
Secondary education	5168	12.22	3860	18.21	976	7.3	30	4.03	302	4.33
University	3299	7.8	2450	11.56	591	4.42	31	4.17	227	3.25
Missing	32324	76.45	13858	65.36	11461	85.77	679	91.26	6326	90.7
Country/region of birth										
Spain	13273	31.39	7534	35.53	3890	29.11	282	37.9	1567	22.47
Western countries[a]	537	1.27	246	1.16	157	1.17	17	2.28	117	1.68
North Africa	502	1.19	306	1.44	152	1.14	0	0	44	0.63
Sub-Saharan Africa	193	0.46	123	0.58	53	0.4	3	0.4	14	0.2
Latin America and the Caribbean	3281	7.76	2129	10.04	719	5.38	53	7.12	380	5.45
Eastern Europe and Central Asia	334	0.79	218	1.03	71	0.53	2	0.27	43	0.62
Asia (not central)[b]	216	0.51	145	0.68	50	0.37	1	0.13	20	0.29
Missing	23947	56.64	10501	49.53	8270	61.89	386	51.88	4790	68.67
Health region of residence										
Alt pirineu i Aran	56	0.13	30	0.14	14	0.1	0	0	12	0.17
Barcelona	35215	83.28	17108	80.69	11566	86.56	708	95.16	5833	83.63
Camp de Tarragona	1655	3.91	930	4.39	390	2.92	10	1.34	325	4.66
Catalunya central	1425	3.37	861	4.06	371	2.78	8	1.08	185	2.65
Girona	2784	6.58	1595	7.52	730	5.46	8	1.08	451	6.47
Lleida	863	2.04	499	2.35	241	1.8	5	0.67	118	1.69
Terres de l'ebre	285	0.67	179	0.84	50	0.37	5	0.67	51	0.73

Western countries[a]: Western Europe, North America, Australia, and New Zealand, Asia (not central)[b]: South-eastern Asia, Southern Asia, and Western Asia.

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract (pages 3 and 6) – COVERED/PERFORMED
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found (page 4) – COVERED/PERFORMED
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported (page 5 and 6) – COVERED/PERFORMED
Objectives	3	State specific objectives, including any prespecified hypotheses (page 6) – COVERED/PERFORMED
Methods		
Study design	4	Present key elements of study design early in the paper (page 6) – COVERED/PERFORMED
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection (page 6 and 7 – COVERED/PERFORMED
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up (page 6-8) – COVERED/PERFORMED
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable (page 6-8, supplementary material tables S1- S4) – COVERED/PERFORMED
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group (page 6-8, supplementary material tables S1- S4) – COVERED/PERFORMED
Bias	9	Describe any efforts to address potential sources of bias (page 6-8 and table S3 in supplementary material) – COVERED/PERFORMED
Study size	10	Explain how the study size was arrived at (page 6 and 7 and table S1 in supplementary material) – COVERED/PERFORMED
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why (page 6-8, supplementary material tables S1 -S4) – COVERED/PERFORMED
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding – COVERED/PERFORMED
		(*b*) Describe any methods used to examine subgroups and interactions – COVERED/PERFORMED
		(*c*) Explain how missing data were addressed – COVERED/PERFORMED
		(*d*) If applicable, explain how loss to follow-up was addressed
		(*e*) Describe any sensitivity analyses
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed – COVERED/PERFORMED
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram

Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (page 6- 8 and table 1 main text, S3 and S5 in supplementary material) – COVERED/PERFORMED
		(b) Indicate number of participants with missing data for each variable of interest (in all tables) – COVERED/PERFORMED
		(c) Summarise follow-up time (eg, average and total amount) (page 6) – COVERED/PERFORMED
Outcome data	15*	Report numbers of outcome events or summary measures over time (page 9-12 and all tables) – COVERED/PERFORMED
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (page 6 and 8 and table 3, for unadjusted estimates) – COVERED/PERFORMED
		(*b*) Report category boundaries when continuous variables were categorized (in all tables) – COVERED/PERFORMED
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives (page 9-12) – COVERED/PERFORMED
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias (page 14) – COVERED/PERFORMED
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence (page 12-15) – COVERED/PERFORMED
Generalisability	21	Discuss the generalisability (external validity) of the study results (page 12-15) – COVERED/PERFORMED
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based (page 17) – COVERED/PERFORMED

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-052817.R2
Article Type:	Original research
Date Submitted by the Author:	30-Oct-2021
Complete List of Authors:	Sentís, Alexis; CEEISCAT; Epiconcept SAS, Epidemiology Montoro-Fernandez, Marcos; CEEISCAT; IGTP Lopez-Corbeto, E; CEEISCAT; CIBERESP Egea-Cortés, Laia; CEEISCAT; IGTP Nomah, DK ; CEEISCAT; Autonomous University of Barcelona, Department of Paediatrics, Obstetrics and Gynecology and Preventive Medicine Díaz, Yesika; CEEISCAT; IGTP Garcia de Olalla, Patricia; Public Health Agency of Barcelona, Epidemiology; CIBERESP Mercuriali, Lilas; Public Health Agency of Barcelona, Epidemiology Borrell, Núria; Public Health Agency of Catalonia, Epidemiological Surveillance and Response to Public Health Emergencies Service in Tarragona Reyes-Urueña, Juliana; CEEISCAT; CIBERESP Casabona, Jordi; ASPC, CEEISCAT; CIBERESP
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	HIV/AIDS, Public health, Research methods, Infectious diseases
Keywords:	Epidemiology < INFECTIOUS DISEASES, HIV & AIDS < INFECTIOUS DISEASES, SEXUAL MEDICINE, Infection control < INFECTIOUS DISEASES, Public health < INFECTIOUS DISEASES, PREVENTIVE MEDICINE

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

Authors: Alexis Sentís,[1,2,3] Marcos Montoro-Fernandez,[1,4] Evelin Lopez-Corbeto,[1,4,5] Laia Egea-Cortés,[1,4] Daniel Kwakye Nomah,[1,6] Yesika Díaz,[1,4] Patrícia Garcia de Olalla,[5,7] Lilas Mercuriali,[7] Núria Borrell,[8] Juliana Reyes-Urueña,[1,4,5] Jordi Casabona,[1,4,5,6] and the Catalan HIV and STI surveillance group*

Affiliations:

[1]Department of Health, Centre of Epidemiological Studies of Sexually Transmitted Disease and AIDS in Catalonia (CEEISCAT), Generalitat of Catalonia, Badalona, Spain

[2]Pompeu Fabra University (UPF), Barcelona, Spain

[3] Epidemiology Department, Epiconcept, Paris, France

[4]Fundació Institut d'Investigació Germans Trias i Pujol (IGTP), Badalona, Spain.

[5]Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain.

[6]Department of Paediatrics, Obstetrics and Gynecology and Preventive Medicine, Universitat Autònoma de Barcelona, Badalona, Spain.

[7]Epidemiology Service. Public Health Agency of Barcelona, Barcelona, Spain.

[8]Epidemiological Surveillance and Response to Public Health Emergencies Service in Tarragona, Agency of Public Health of Catalonia, Generalitat of Catalonia, Tarragona, Spain

*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 Mercuriali, 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 Martinez, 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 Mendioroz, 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).

Correspondence to:

Alexis Sentís Fuster

Address: Fundació Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol (IGTP), Edifici Muntanya, Carretera de Can Ruti, Camí de les Escoles s/n, 08916 Badalona (Spain)

E-mail address: alexissentis@gmail.com,

Journal: BMJ Open

- Article type**:** Original article
- Word count: 3992
- Figures and tables: 1 figure, 4 tables (max: 5 tables and/or figures)
- References: 30
- Supplementary material: 5 tables

Key words: Epidemiology, infection control, public health, HIV & AIDS, sexual medicine, preventive medicine

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 (LGV) cases among 34,600 individuals who 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 2017-2019 (*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 odds ratio [aOR] 23.69; 95% confidence interval [CI] 16.67–35.13), age 30–39 years (versus <20 years, aOR 18.58; 95% CI 8.56–52.13), having 5–7 STI episodes (versus 1 episode, aOR 5.96; 95% CI 4.26–8.24) and living in urban areas (aOR 1.32; 95% CI 1.04–1.69). Living in the most deprived BHAs (aOR 0.60; 95% CI 0.50–0.72) 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 MSM in urban and less deprived areas showed higher rates of incidence, multiple STI episodes, and HIV coinfection.

Conclusions: We recommend socio-epidemiological identification and characterization of STI clusters and factors associated with HIV coinfection to identify at-risk populations at a health area level to design effective interventions.

Word count: 300 (max: 300 words)

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 characterize the re-emergence of STIs, perform socio-epidemiological clustering and reveal factors associated with HIV co-infection.
- To our knowledge, this is the first study to apply the k-means clustering methodology to identify and characterize 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 socio-demographic and lifestyle characteristics such as education level, sexual preference and country of birth. Nonetheless, our findings are consistent with previous analyses.

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.[1,2] 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.[3–6] 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.[7,8] 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.[9] 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.[7,9] 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, socio-cultural shifts in society, and the effects of tourism and globalization.[10]

STIs and HIV infections are overlapping epidemics, which, besides from biological synergies, are largely driven by socio-economic 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.[11,12] Some studies have described social determinants of health, discrimination and inequalities as the main factors associated with the spatiotemporal clustering of STI cases.[13,14] 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 sociological and epidemiological factors.[15] The socio-epidemiological characterization 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,[16] 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 supplementary table S1). Case definitions for surveillance reporting are standardized according to the European Union definitions established by the European Centre for Disease Prevention and Control (online supplementary table S2).[17,18]

Analysis variables

We extracted data around epidemiological, socio-demographic and clinical variables as detailed in online supplementary 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, Catalan HIV/STI Registry surveillance team performs duplicate checks 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 anonymized 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 5,000–25,000 people. The socio-economic level of the BHAs were 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 categorized in quintiles, with the first quintile being the least deprived.[19]

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 supplementary Table S3). As information regarding treatment response was not available, episodes occurring outside of the specific timeframes 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 socio-demographic characteristics. Specifically, the k-means clustering methodology is an unsupervised machine learning approach that seeks to group heterogenous units (in our case BHAs) into clusters based on similarities in characteristics (variable values and categorical distribution among the BHA).[15] 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 intra-cluster 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 dataset. 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 summarize epidemiological, clinical, socio-demographic and geographical variables for the total confirmed STI cases and by STI clusters. Continuous variables were summarised as median (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 supplementary Table S4). Incidence trends were analysed using the χ^2 test for linear trend. For identifying and building clusters in k-means clustering, STI incidence rates were described per 1,000 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 (version 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 de-identified 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 9,687 in 2017 to 18,872 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 ≥30 years). Indeed, women under 30 years of age presented the highest decrease in both number of cases, with an increase of 155.8% vs 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) vs 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% 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, 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 supplementary 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 greater among males (aOR 23.69; 95% CI 16.67–35.13 compared with females) and in urban BHAs (aOR 1.32; 95% CI 1.04–1.69 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 three years of the study period. This fact and having 5,773 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 versus 160.6 per 100,000 population), but represented the majority of all reported STI cases (55.7%). STI-infected individuals in Cluster 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% versus 11.0%) and had a higher proportion of heterosexual men and women (approximately 12% higher) and chlamydia cases (61.7% versus 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 100,000 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 4,359 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 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. Further, 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.[3–6] The rise in STI cases has been partially attributed to enhancement of surveillance systems and the introduction of improved diagnostic tools in recent years.[10] 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.[20–22]

Chlamydia has been reported more frequently in WSM, while syphilis, gonorrhoea and LGV were more common in MSW and MSM.[3–9] 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 of STIs in Catalonia in 2007–2015,[12] 2012-2017[23] and 2018-2019,[9] 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 socio-economic status and education levels.[24]

Consistent with previous data,[12] 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.[25] The World Health Organization and other public health agencies have emphasized 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 characterization of populations at increased risk of STIs.[10,25] Socio-demographic and socio-economic are increasingly being established as more important risk factors of STI acquisition than individual behaviours, particularly among women from disadvantaged groups. [26,27]

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 heterogenous variables such as socio-demographic, clinical, cognitive, vital signs and laboratory parameters.[28] More recently, its potential to monitor and group trends of SARS-CoV-2 prevalence by magnitude (higher, medium and lower) at a regional level in Italy has been described.[29] Identification of these 'clusters of characteristics' may be useful, in their specific context, to better detect and characterize 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.[30] 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 socio-demographic 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.[3–6,9,12,24] The age category above 60 years old may contribute to residual cofounding 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 socio-economic 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 characterization 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 not only an issue of the health sector, 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, not only among the MSM population, but 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 prevalence as the key monitoring variable.

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.

Competing interests

All authors declare no potential conflicts of interest relevant to this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Contributors

AS conceptualized and designed the study. MMF cleaned the database, MMF, LEC and YD performed the statistical and cluster analysis. AS, ELC and DN reviewed scientific literature, AS, JRU, and JC drafted the manuscript and AS, ELC, PGO, LM, NB, JRU and JC interpreted the results. All authors critically reviewed the manuscript and approved the final version to be published.

Data sharing statement

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

REFERENCES

1. 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;97:548-562P.

2. WHO Department of Reproductive Health and Research. Report on global sexually transmitted infection surveillance, 2018. Available at: https://apps.who.int/iris/bitstream/handle/10665/277258/9789241565691-eng.pdf?ua=1 [Accessed 8 January 8 2021].

3. European Centre for Disease Prevention and Control. Chlamydia infection. In: ECDC. Annual epidemiological report for 2018. Stockholm: ECDC; 2020. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/AER-for-2018-STI-chlamydia.pdf [Accessed 7 August 2021].

4. European Centre for Disease Prevention and Control. Gonorrhoea. In: ECDC. Annual epidemiological report for 2018. Stockholm: ECDC; 2020. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/gonorrhoea-annual-epidemiological-report-2018.pdf [Accessed 7 August 2021].

5. European Centre for Disease Prevention and Control. Syphilis. In: ECDC. Annual epidemiological report for 2018. Stockholm: ECDC; 2020. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/syphilis-aer-2018.pdf [Accessed 7 August 2021].

6. European Centre for Disease Prevention and Control. Lymphogranuloma venereum. In: ECDC. Annual epidemiological report for 2018. Stockholm: ECDC; 2020. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/lymphogranuloma-venereum-aer-2018.pdf [Accessed 7 August 2021].

7. Centro Nacional de Epidemiología. Unidad de vigilancia del VIH y conductas de riesgo. Vigilancia epidemiológica 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, Calidad e Innovación; 2019. Available at: https://www.mscbs.gob.es/ciudadanos/enfLesiones/enfTransmisibles/sida/vigilancia/Vigilancia_ITS_1995_2017_def.pdf [Accessed 27 March 2021]

8. Área de vigilancia del VIH y conductas de riesgo. Vigilancia epidemiológica de las infecciones de transmisión sexual, 1995-2010. Madrid: Centro Nacional de Epidemiología; 2012. Available at:

https://www.mscbs.gob.es/ciudadanos/enfLesiones/enfTransmisibles/sida/vigilancia/VigilanciaITS1995_2010.pdf [Accessed 27 March 2021].

9. Centre d'Estudis Epidemiològics sobres 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: CEEISACT; 2020. Available at: https://canalsalut.gencat.cat/web/.content/_A-Z/S/sida/enllasos/anual_ITS.pdf [Accessed 27 March 2021].

10. Taylor MM, Wi TEC. Transforming and integrating STI surveillance to enhance global advocacy and investment in STI control. *J Int AIDS Soc* 2019;22:25361.

11. Centers for Disease Control and Prevention. HIV/AIDS & STDs.Available at: https://www.cdc.gov/std/hiv/default.htm. [Accessed 22 February 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; 2015. Available at: https://scientiasalut.gencat.cat/bitstream/handle/11351/3418/informe_SIVES_2015_informe_epidemiologic_CEEISCAT_2015.pdf.pdf?sequence=1&isAllowed=y [Accessed 27 March 2021].

17. Notifiable diseases and epidemic outbreaks. Department of Health, Public Health Agency of Catalonia, Generalitat de Catalunya; 2016. Available at: https://canalsalut.gencat.cat/ca/professionals/vigilancia-epidemiologica/malalties-de-declaracio-obligatoria-i-brots-epidemics/ [Accessed 20 June 2021].

18. European Centre for Disease Prevention and Control. EU case definitions; 2018. Available at: https://www.ecdc.europa.eu/en/surveillance-and-disease-data/eu-case-definitions. [Accessed 7 August 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 at: http://observatorisalut.gencat.cat/ca/observatori-desigualtats-salut/indicador_socioeconomic_2015/ [Accessed 6 August 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. 2019;18:1–11.

22. Jansen K, Steffen G, Potthoff A, *et al.* STI in times of PrEP: High prevalence of chlamydia, gonorrhea, and mycoplasma at different anatomic sites in men who have sex with men in Germany. *BMC Infect Dis* 2020;20:110.

23. Vives N, Olalla PG de, 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, Gonorrhea 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. *Canada Commun Dis Rep* 2019;45:63–6.

26. Hill A V., 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 Heal* 2019;64:319–26.

27. Jagadesh S, Combe M, Couppié P *et al.* Mapping priority neighborhoods: A novel approach to cluster identication in HIV/AIDS population. *Res Sq* 2020 [doi: 10.21203/rs.3.rs-17843/v1].

28. de la Fuente-Tomas L, Arranz B, Safont G, *et al.* Classification of patients with bipolar disorder using k-means clustering. *PLoS One* 2019;14.

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;17:1–14.

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.

FIGURE LEGEND

Figure 1. Incidence rates (per 1,000 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 visualization of Cluster C. From a total of 373 Catalan BHA five (Garraf rural, Polinyà-Sentmenat, Ribes-Olivella. Roquetes-Canyelles, Viladecans 3) were excluded from the K-means clustering analysis because their delimitations and populations changed during the study period.

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

Characteristic	All STIs	Chlamydia	Gonorrhoea	Syphilis	LGV
Total cases, n (%)	42,283 (100)	21,202 (50.1)	13,362 (31.6)	6,975 (16.5)	744 (1.8)
Sex, n (%)					
Female	16,676 (39.4)	13,125 (61.9)	2,667 (20.0)	875 (12.5)	9 (1.2)
Male	25,607 (60.6)	8,077 (38.1)	10,695 (80.0)	6,100 (87.5)	735 (98.8)
Age, n (%)					
<20 years	4,438 (10.5)	3,311 (15.6)	984 (7.4)	137 (2.0)	6 (0.8)
20–29 years	17,691 (41.8)	10,707 (50.5)	5,361 (40.1)	1,462 (21.0)	161 (21.6)
30–39 years	11,102 (26.3)	4,454 (21.0)	4,116 (30.8)	2,242 (32.1)	290 (39.0)
40–49 years	6,092 (14.4)	2,087 (9.8)	2,032 (15.2)	1,757 (25.2)	216 (29.0)
50–59 years	2,037 (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)
Sexual preference, n (%)					
MSM*	3,270 (7.7)	785 (3.7)	1,321 (9.9)	993 (14.2)	171 (23.0)
MSW	3,149 (7.5)	1,863 (8.8)	1,040 (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	8,189 (19.4)	7,034 (33.2)	966 (7.2)	186 (2.7)	3 (0.4)
Missing (male)	19,188 (45.4)	5,429 (25.6)	8,334 (62.4)	4,864 (69.7)	561 (75.4)
Missing (female)	8,072 (19.1)	5,756 (27.2)	1,632 (12.2)	679 (9.7)	5 (0.7)
STI reinfection, n (%)	4,558 (10.8)	1,418 (6.7)	2,098 (15.7)	955 (13.7)	87 (11.7)
HIV coinfection, n (%)	2,443 (5.8)	467 (2.2)	897 (6.7)	893 (12.8)	186 (25.0)
Deprivation index‡					
First quintile (least deprived)	10,271 (24.3)	5,185 (24.5)	3,040 (22.8)	1,757 (25.2)	289 (38.8)
Second quintile	7,465 (17.7)	4,328 (20.4)	2,012 (15.1)	1,037 (14.9)	88 (11.8)
Third quintile	4,859 (11.5)	2,763 (13.0)	1,332 (10.0)	716 (10.3)	48 (6.5)
Fourth quintile	5,703 (13.5)	3,217 (15.2)	1,578 (11.8)	827 (11.9)	81 (10.9)
Fifth quintile	7,689 (18.2)	4,319 (20.4)	2,211 (16.6)	1,079 (15.5)	80 (10.8)
Missing	6,296 (14.9)	1,390 (6.6)	3,189 (23.9)	1,559 (22.4)	158 (21.2)
Health region of residence, n (%)					
Barcelona	35,215 (83.3)	17,108 (80.7)	11,566 (86.6)	5,833 (83.6)	708 (95.2)
Other regions	7,068 (16.7)	4,094 (19.3)	1,796 (13.4)	1,142 (16.4)	36 (4.8)
BHA setting					
Rural	4,193 (9.9)	2,614 (12.3)	1,039 (7.8)	516 (7.4)	24 (3.2)
Urban	29,969 (70.9)	16,347 (77.1)	8,566 (64.1)	4,516 (64.8)	540 (72.6)
Missing	8,121 (19.2)	2,241 (10.6)	3,757 (28.1)	1,943 (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.

‡1st quintile (31.52%), 2nd quintile (40.09%), 3rd quintile (46.27%), 4th quintile (53.98%), 5th 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.

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	
Total							
Total	9,687	128.2	13,724	180.6	18,872	245.9	<0.001
Sex							
Female[a]	3,362	87.4	5,503	142.2	7,811	200.0	<0.001
Male[b]	6,325	170.5	8,221	220.4	11,061	293.4	<0.001
Age, years							
<30[c]	4,607	198.4	7,181	306.9	10,341	436.5	<0.001
≥30[d]	5,080	97.1	6,543	124.4	8,531	160.8	<0.001
Sex and age							
Male <30 years[f]	2,411	202.9	3,347	279.5	4,668	384.0	<0.001
Male ≥30 years[g]	3,914	155.2	4,874	192.4	6,393	250.3	<0.001
Female <30 years[h]	2,196	193.6	3,834	335.7	5,673	491.9	<0.001
Female ≥30 years[I]	1,166	43.0	1,669	61.2	2,138	77.7	<0.001
STI type (total Catalonia[j])							
Chlamydia	3,562	47.1	7,240	95.3	10,400	135.5	<0.001
Gonorrhoea	3,492	46.2	4,088	53.8	5,782	75.3	<0.001
Syphilis	2,430	32.2	2,175	28.6	2,370	30.9	0.1572
LGV	203	2.7	221	2.9	320	4.2	<0.001
Health region of residence							
Alt pirineu i Aran[k]	10	13.9	13	18.1	33	45.7	<0.001
Barcelona[L]	8,205	165.0	11,475	229.5	15,535	307.8	<0.001
Camp de Tarragona[m]	294	49.1	491	81.3	870	142.2	<0.001
Catalunya Central[n]	358	69.4	484	93.1	583	110.7	<0.001
Girona[o]	564	65.7	893	103.2	1,327	151.5	<0.001
Lleida[p]	212	58.9	247	68.5	404	111.5	<0.001
Terres de l'Ebre[q]	44	24.5	121	67.7	120	67.2	<0.001
BHA setting[a]							
Rural	827	NA	1,375	NA	1,991	NA	NA
Urban	6,475	NA	9,868	NA	13,626	NA	NA
Missing	2,385	NA	2,481	NA	3,255	NA	NA

BHA, Basic Health Area; LGV, lymphogranuloma venereum; STI, sexually transmitted infection. Data source for denominators (a-q): the Statistical Institute of Catalonia (IDESCAT): online supplementary Table S4.

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=1,376)	OR	95% CI	aOR	95% CI
Sex						
Female	14,938	29	1 (ref)		1 (ref)	
Male	19,662	1,347	37.81	26.69–55.93	23.69	16.67–35.13
Age group, years						
<20	3,696	5	1 (ref)		1 (ref)	
20-29	14,826	328	16.70	7.70–46.83	8.33	3.82–23.40
30-39	8,704	595	54.17	25.05–151.57	18.58	8.56–52.13
40-49	4,759	339	56.62	26.09–158.78	17.66	8.10–49.65
50-59	1,748	89	39.60	17.80–112.58	13.06	5.84–37.24
>60	867	20	17.43	7.04–52.50	6.98	2.80–21.09
Deprivation index*						
First quintile (least deprived)	7,679	501	1 (ref)		1 (ref)	
Second quintile	6,098	210	0.51	0.43–0.60	0.70	0.59–0.83
Third quintile	4,163	109	0.38	0.31–0.47	0.63	0.50–0.78
Fourth quintile	4,663	186	0.60	0.50–0.71	0.83	0.69–1.00
Fifth quintile	6,347	175	0.41	0.34–0.48	0.60	0.50–0.72
Missing	5,650	195	0.51	0.43–0.60	0.51	0.39–0.67
STI episodes (total), n						
1	29,104	791	1 (ref)		1 (ref)	
2–4	5,304	529	3.96	3.54–4.44	2.69	2.39–3.03
5–7	192	56	14.74	10.64–20.16	5.96	4.26–8.24
BHA setting						
Rural	3,699	81	1 (ref)		1 (ref)	
Urban	23,812	1,023	2	1.61–2.54	1.32	1.04–1.69

*1st quintile (31.52%), 2nd quintile (40.09%), 3rd quintile (46.27%), 4th quintile (53.98%), 5th quintile (100%).

aOR, adjusted odds ratio; BHA, Basic Health Area; CI, confidence interval; LGV, lymphogranuloma venereum; MSM, men who have sex with men; MSW, men who have sex with women; OR, odds ratio; STI, sexually transmitted infection; WSM, women who have sex with men; WSW, women who have sex with women.

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

Characteristic	Cluster A	Cluster B	Cluster C	Total*
Demographics				
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
Reported STI cases, n (%)				
Total	11,527 (32.2)	19,945 (55.7)	4,359 (12.2)	35,831 (100)
Sex				
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)
Country of birth				
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)
Sexual preference				
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)
STI type				
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)
STI incidence rate categories of BHAs				
First quintile (2.4 per 1,000)	710 (6.2)	1820 (9.1)	0	2530 (7.1)
Second quintile (3.6 per 1,000)	2136 (18.5)	3359 (16.8)	0	5495 (15.3)
Third quintile (5.2 per 1,000)	1377 (12.0)	6588 (33.0)	0	7965 (22.2)
Fourth quintile (9.8 per 1,000)	5688 (49.4)	7508 (37.6)	0	13196 (36.8)
Fifth quintile (42.8 per 1,000)	1616 (14.0)	670 (3.4)	4359 (100)	6645 (18.6)

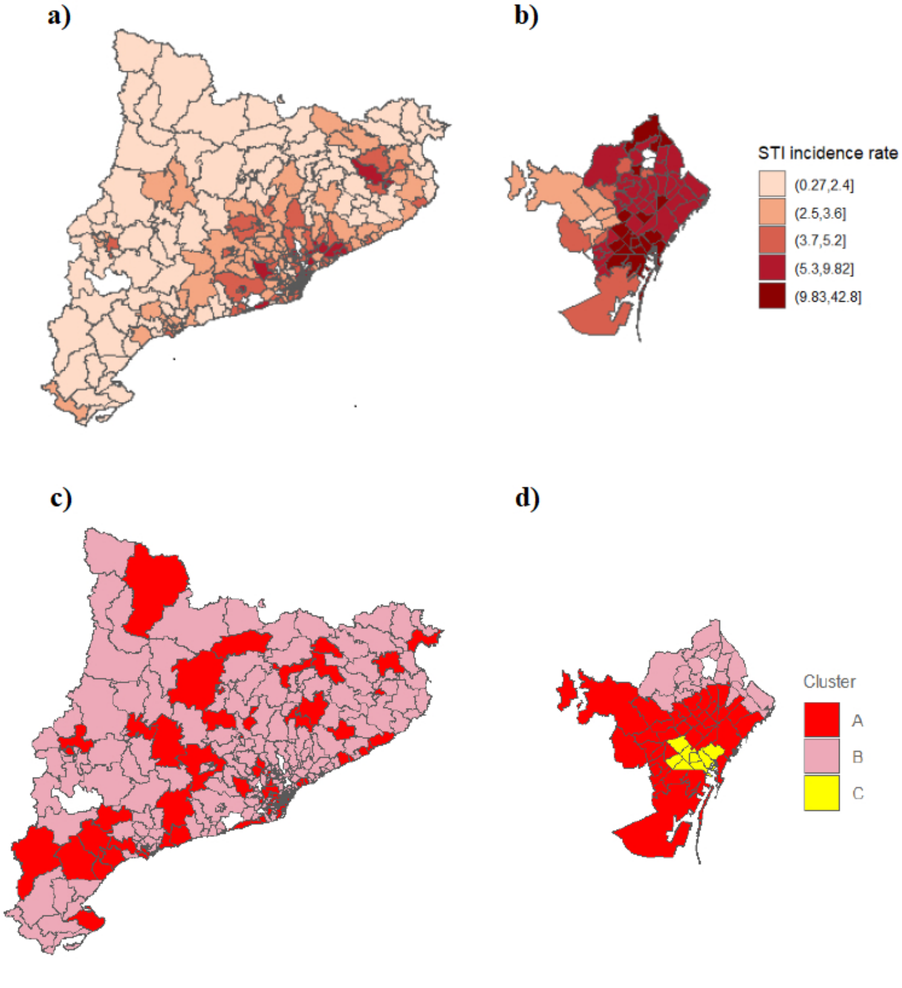
BHA setting				
Rural	797 (6.9)	3,158 (15.8)	0	3,955 (11.0)
Urban	9,461 (82.1)	15,787 (79.2)	4,359 (100)	29,607 (82.6)
Missing	1,269 (11.0)	1,000 (5.0)	0	2,269 (6.3)

*Of the 373 Catalan BHAs, five (Garraf rural, Polinyà-Sentmenat, Ribes-Olivella, Roquetes-Canyelles, 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; 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.



218x238mm (96 x 96 DPI)

SUPPLEMENTARY MATERIAL

Table S1. Epidemiological Repository of Catalonia (REC, in Catalan), an electronic registry used by the Epidemiological Surveillance Network of Catalonia (XVEC, in Catalan)

As instituted by Law 203/2015 (September 15, 2015), personnel from the Epidemiological Surveillance Network of Catalonia (XVEC) manage the mandatory declaration of diseases and epidemic outbreaks. Along with the notification, healthcare professionals enclose a questionnaire that describes epidemiological, behavioural, clinical, and geographical parameters. The notification comes from two main sources. The first is the Mandatory Declaration of Disease (MDO) system, where a healthcare professional reports a suspected or confirmed case using established case definitions. The notification procedure is done electronically or, alternatively, by means of the individualized notice form on paper. In compliance with article 13 of law 67/2010 (25 May 2010) of the Health Department of Government of Catalonia, nominal notification of syphilis, gonorrhoea, and LGV have been reported to the MDO since 2006, chlamydia since 2015 and congenital syphilis since 1997. The second source of notification is the Microbiological Notification System of Catalonia (SNMC), which collects microbiological information on selected diseases. Notifications on chlamydia and gonorrhoea are also reported through the SNMC. Notification of new HIV infections was done on a voluntary basis between 2001 and 2009, and mandatory and nominal since 2010.

Table S2. STI case definitions from the Public Health Agency of Catalonia, Department of Health, Generalitat de Catalunya.

CHLAMYDIA: **1) Laboratory criteria for diagnosis:** Isolation of *Chlamydia trachomatis* by culture in a sample of the genitourinary tract, anal or conjunctiva, or clinical sample; or demonstration of *C. trachomatis* by detection of specific antigens or by direct immunofluorescence (DFA) in a clinical sample; or detection of specific genomic fragments of *C. trachomatis* in a clinical specimen. **2) Confirmed case:** Person with compatible laboratory criteria. **3) Probable case:** Person with clinically compatible criteria, especially if it is epidemiologically related.

GONORRHOEA: **1) Laboratory criteria for diagnosis:** Isolation by culture of *Neisseria gonorrhoeae* in a clinical specimen, or detection of specific genomic fragments of *N. gonorrhoeae* in a clinical specimen, or microscopic detection of gram-negative intracellular diplococci in urethral exudates in men. **2) Confirmed case:** Person with compatible laboratory criteria. **3) Probable case:** Person with clinically compatible criteria, especially if it is epidemiologically related.

SYPHILIS: **1) Laboratory criteria for diagnosis**: Demonstration of *Treponema pallidum* by dark field microscopy, by direct immunofluorescence (DFA), of genomic fragments, in lesion secretions. Detection of antibodies against *T. pallidum* by specific tests (TPHA, TPPA or EIA) and, in addition, one of the following methods: FTA-ABS, EIA immunotransference, non-specific reactive serological test (VDRL, RPR), detection of IgM antibodies -TP. **2) Confirmed case**: Person with compatible laboratory criteria. **3) Probable case:** Person with clinically compatible criteria, especially if it is epidemiologically related.

LGV: **1) Laboratory criteria for diagnosis:** Detection of specific genomic fragments of *C. trachomatis* in a clinical sample, and in addition Identification of serovar L1, L2 or L3. **2) Confirmed case:** Person with compatible laboratory criteria. **3) Probable case:** Person with clinically compatible criteria, especially if it is epidemiologically related.

Notifiable diseases and epidemic outbreaks. Department of Health, Public Health Agency of Catalonia, Generalitat de Catalunya. Available at: https://canalsalut.gencat.cat/ca/professionals/vigilancia-epidemiologica/malalties-de-declaracio-obligatoria-i-brots-epidemics/ ((accessed 20 Jun 2021).

Table S3. Details on study variables.

The socio-demographic variables used in our analysis were sex, age at notification, educational level, deprivation index and country of birth. We used a basic health area (ABS) deprivation index calculated by the Agency for Health Quality and Assessment of Catalonia (AQUAS), attributed to each patient according to their address of residence (categorized in quintiles, first quintile for the ABS with lower deprivation index)[a]. We extracted the classification of ABS as urban or rural (ABS urbanicity) from another deprivation index at ABS level provided by the Primary Health Care Information Systems (SISAP), MEDEA index[b]. Country of birth were categorized by regions adapting for the study those used by WHO. We categorized sexual preference separately for men and women as follows: Men (two groups): MSM (include men who have sex with men, bisexual men and transgender men) and men who have sex with women only (MSW); Women (two groups): WSW (includes women who have sex with women, bisexual women, transgender men) and women who have sex only with men (WSM). Some variables from the epidemiological questionnaire showed high percentages of missing values such are education level, country of birth, and sexual preference (76%, 57%, and 64% respectively, see table 1 and S4). The clinical variables were reinfections, multiple STI episodes (when same persona had more than one during the study period), and coinfection with HIV. Reinfection was defined as more than one episode of the same specific STI during all the study follow-up, but defined differently for each STI, depending on the number of days between successive episodes after the first infection in the same individual; more than 364 for syphilis (although definitive criteria for cure or failure have not been well established yet) and 119 days for gonorrhoea, chlamydia and LGV, respectively[c]. As a geographical variable, we categorized people based on the seven Catalan health regions of their ABS of residence: Alt Pirineu and Aran, Barcelona, Camp de Tarragona, Catalunya central, Girona, Lleida and Terres de l'ebre.

[a] 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.http://observatorisalut.gencat.cat/ca/observatori-desigualtats-salut/indicador_socioeconomic_2015/ (accessed 6 Aug 2020).

[b] Domínguez-Berjón MF, Borrell C, Cano-Serral G, et al. Construcción de un índice de privación a partir de datos censales en grandes ciudades españolas (Proyecto MEDEA). Gac Sanit 2008;22:179–87. doi:10.1157/13123961

[c] CDC - STD Treatment. https://www.cdc.gov/std/treatment/default.htm (accessed 14 Feb 2021).

Table S4. Denominators for STI incidence rates calculations.

Data source for denominators (a-q) in Table 2 (main text): the Statistical Institute of Catalonia (IDESCAT) [data provided by IDESCAT on 23 June 2020]:

[a]Total female population in Catalonia: 3,845,630 in 2017, 3,869,739 in 2018, and 3,905,094 in 2019.

[b]Total male population in Catalonia: 3,710,200 in 2017, 3,730,326 in 2018, and 3,770,123 in 2019.

[c]Total population <30 yrs in Catalonia: 2,322,227 in 2017, 2,339,673 in 2018, and 2,368,830 in 2019.

[d]Total population ≥ 30 yrs in Catalonia: 5,233,603 in 2017, 5,260,392 in 2018, and 5,306,387 in 2019.

[e]Total male population<30 yrs in Catalonia: 1,187,850 in 2017, 1,197,499 in 2018, and 1,215,583 in 2019.

[f]Total male population≥ 30 yrs in Catalonia: 2,522,350 in 2017, 2,532,827 in 2018, and 2,554,540 in 2019.

[g]Total female population<30yrs in Catalonia: 1,134,377 in 2017, 1,142,174 in 2018, and 1,153,247 in 2019.

[h]Total female population≥ 30yrs in Catalonia: 2,711,253 in 2017, 2,727,565 in 2018, and 2,751,847 in 2019.

[I]Total population in Catalonia: 7,555,830 in 2017, 7,600,065 in 2018, and 7,675,217 in 2019. Used as a denominator to compute STI rates, total and by disease, and rates according ABS urbanicity category.

[j]Total population in Alt pirineu i Aran: 71,958 in 2017, 71,888 in 2018, and 72,276 in 2019.

[k]Total population in Barcelona: 4,972,179 in 2017, 5,000,125 in 2018, and 5,047,597 in 2019.

[L]Total population in Camp de Tarragona: 598,683 in 2017, 603,743 in 2018, and 611,950 in 2019.

[m]Total population in Catalunya Central: 515,578 in 2017, 519,819 in 2018, and 526,544 in 2019.

[n]Total population in Girona: 857,877 in 2017, 865,282 in 2018, and 875,722 in 2019.

[o]Total population in Lleida: 359,729 in 2017, 360,497 in 2018, and 362,428 in 2019.

[p]Total population in Terres de l'Ebre: 179,826 in 2017, 178,711 in 2018, and 178,700 in 2019.

[q]NA=Not available denominators.

Table S5. Distribution of epidemiological characteristics in cases of chlamydia, gonorrhoea, syphilis or lymphogranuloma venerum (LGV) in Catalonia, 2017–2019 (N= 42,283).

	All STI (N=42,283)		Chlamydia (N=21,202)		Gonorrhoea (N=13,362)		Syphilis (N=6,975)		LGV (N =744)	
	N	%	N	%	N	%	N	%	N	%
Education										
Primary school or less	1492	3.53	1034	4.88	334	2.5	4	0.54	120	1.72
Secondary education	5168	12.22	3860	18.21	976	7.3	30	4.03	302	4.33
University	3299	7.8	2450	11.56	591	4.42	31	4.17	227	3.25
Missing	32324	76.45	13858	65.36	11461	85.77	679	91.26	6326	90.7
Country/region of birth										
Spain	13273	31.39	7534	35.53	3890	29.11	282	37.9	1567	22.47
Western countries[a]	537	1.27	246	1.16	157	1.17	17	2.28	117	1.68
North Africa	502	1.19	306	1.44	152	1.14	0	0	44	0.63
Sub-Saharan Africa	193	0.46	123	0.58	53	0.4	3	0.4	14	0.2
Latin America and the Caribbean	3281	7.76	2129	10.04	719	5.38	53	7.12	380	5.45
Eastern Europe and Central Asia	334	0.79	218	1.03	71	0.53	2	0.27	43	0.62
Asia (not central)[b]	216	0.51	145	0.68	50	0.37	1	0.13	20	0.29
Missing	23947	56.64	10501	49.53	8270	61.89	386	51.88	4790	68.67
Health region of residence										
Alt pirineu i Aran	56	0.13	30	0.14	14	0.1	0	0	12	0.17
Barcelona	35215	83.28	17108	80.69	11566	86.56	708	95.16	5833	83.63
Camp de Tarragona	1655	3.91	930	4.39	390	2.92	10	1.34	325	4.66
Catalunya central	1425	3.37	861	4.06	371	2.78	8	1.08	185	2.65
Girona	2784	6.58	1595	7.52	730	5.46	8	1.08	451	6.47
Lleida	863	2.04	499	2.35	241	1.8	5	0.67	118	1.69
Terres de l'ebre	285	0.67	179	0.84	50	0.37	5	0.67	51	0.73

Western countries[a]: Western Europe, North America, Australia, and New Zealand, Asia (not central)[b]: South-eastern Asia, Southern Asia, and Western Asia.

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract (pages 3 and 6) – COVERED/PERFORMED
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found (page 4) – COVERED/PERFORMED
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported (page 5 and 6) – COVERED/PERFORMED
Objectives	3	State specific objectives, including any prespecified hypotheses (page 6) – COVERED/PERFORMED
Methods		
Study design	4	Present key elements of study design early in the paper (page 6) – COVERED/PERFORMED
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection (page 6 and 7 – COVERED/PERFORMED
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up (page 6-8) – COVERED/PERFORMED
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable (page 6-8, supplementary material tables S1- S4) – COVERED/PERFORMED
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group (page 6-8, supplementary material tables S1- S4) – COVERED/PERFORMED
Bias	9	Describe any efforts to address potential sources of bias (page 6-8 and table S3 in supplementary material) – COVERED/PERFORMED
Study size	10	Explain how the study size was arrived at (page 6 and 7 and table S1 in supplementary material) – COVERED/PERFORMED
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why (page 6-8, supplementary material tables S1 -S4) – COVERED/PERFORMED
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding – COVERED/PERFORMED
		(*b*) Describe any methods used to examine subgroups and interactions – COVERED/PERFORMED
		(*c*) Explain how missing data were addressed – COVERED/PERFORMED
		(*d*) If applicable, explain how loss to follow-up was addressed
		(*e*) Describe any sensitivity analyses
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed – COVERED/PERFORMED
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram

Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (page 6- 8 and table 1 main text, S3 and S5 in supplementary material) – COVERED/PERFORMED
		(b) Indicate number of participants with missing data for each variable of interest (in all tables) – COVERED/PERFORMED
		(c) Summarise follow-up time (eg, average and total amount) (page 6) – COVERED/PERFORMED
Outcome data	15*	Report numbers of outcome events or summary measures over time (page 9-12 and all tables) – COVERED/PERFORMED
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (page 6 and 8 and table 3, for unadjusted estimates) – COVERED/PERFORMED
		(*b*) Report category boundaries when continuous variables were categorized (in all tables) – COVERED/PERFORMED
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives (page 9-12) – COVERED/PERFORMED
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias (page 14) – COVERED/PERFORMED
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence (page 12-15) – COVERED/PERFORMED
Generalisability	21	Discuss the generalisability (external validity) of the study results (page 12-15) – COVERED/PERFORMED
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based (page 17) – COVERED/PERFORMED

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.