Prevalence of anatomically specific infections with *Chlamydia trachomatis* among men who have sex with men in China: protocol for a nationwide cross-sectional study as part of Disease Burden Surveillance of Infections with *Chlamydia* (DBSIC)

Ting-Ting Jiang, Yan Han, Ying Zhou, Ning-Xiao Cao, Mei-Qin Shi, Yue Ping Yin, Xiang-Sheng Chen

**ABSTRACT**

**Introduction** *Chlamydia trachomatis* (CT) causes the most prevalent bacterial sexually transmitted infection in the world, and men who have sex with men (MSM) are considered as a high-risk population for this infection. Data regarding the prevalence of CT infection in China are limited and fragmented. In this study, we aim to determine at the national level the anatomically specific CT prevalence and genotype distribution of CT strains among MSM, recruiting from the venues where MSM commonly seek sexual partners.

**Methods and analysis** The study will be a nationwide cross-sectional survey to estimate the prevalence of chlamydial infections among MSM who meet the inclusion criteria over a period of 6 months (May–October 2022). A total of 2429 participants will be recruited from the venues where MSM most often seek sex partners in 14 cities of all 7 geographical regions in China. A mobile phone app-based anonymous self-administered questionnaire will be used to collect sociodemographic and behavioural data, and specimens of urine, anorectal and pharyngeal swabs will be collected for identifying the infections of CT and *Neisseria gonorrhoeae* and genotypes of CT. The data will be analysed using the IBM SPSS program V20

**Ethics and dissemination** The study protocol has been approved by the Medical Ethics Committee of the Chinese Academy of Medical Sciences Institute of Dermatology and the National Center for STD Control on 9 October 2021 (approval number 2021-KY-037). Any minimal changes will be reported in the final manuscript to ensure the integrity of the study. The dataset of control programmes and used as health education materials to disseminate to the community. The dataset will be deposited in a public repository.

**Trial registration number** ChiCTR2100052869.

**Strengths and limitations of this study**

- This will be the first nationwide study conducted in China to estimate prevalence of chlamydial infections in different anatomical sites and investigate genotype distribution of chlamydia strains among men who have sex with men (MSM).
- The study includes a study sample of the MSM population from 14 cities of all 7 geographical regions in China.
- A mobile phone app-based anonymous self-administered questionnaire to collect the sociodemographic and behavioural data is facilitated to improve data quality.
- Due to operational feasibility, we are not able to include all kinds of the venues where MSM seek sexual partners into the study or recruit those MSM who may not attend any specific venues, potentially compromising the representativeness of the study population.

**INTRODUCTION**

*Chlamydia trachomatis* (CT) has caused one of the most prevalent bacterial sexually transmitted infections (STIs) globally, representing 33.8% and 43.6% of incident and prevalent cases of the four curable STIs (chlamydia, gonorrhea, trichomoniasis and syphilis) in women and men, aged 15–49 years, in 2016. The group of males who have sexual contacts with other males has been categorised as men who have sex with men (MSM) in the context of their increased risk of STIs. Urethra, rectum and pharynx are the common sites of STIs among MSM due to unprotected oral or anal sexual behaviours and may serve as
infection prevalence, particularly in the anorectal sites.6 However, the national CT infection rates in China are still not clearly known because there are neither comprehensive CT surveillance systems so far nor epidemiological data that are considered representative at national levels. Therefore, the primary objective of this study is to estimate the national prevalence of CT infections and identify the predictors of the infections among the MSM population in China. The key secondary objectives are to estimate the prevalence of Neisseria gonorrhoeae (NG) infections, analyse the genotype distribution of CT strains obtained from different anatomical (urethral, anorectal and pharyngeal) sites, identify the genotypes related to lymphogranuloma venereum (LGV) and compare the genotype distribution across the geographical regions in China with those of the strains genotyped elsewhere in the world.

METHODS AND ANALYSIS

Study design
This will be a multisite cross-sectional study over a period of 6 months starting from May 2022, and the estimated date to complete the recruitment of the last participant is prior to 31 October 2022.

Sampling design and setting
In China, the distribution of MSM population is strongly associated with geographical regions with different cultural and economic development. Administratively, China consists of 23 provinces, 4 municipalities and 5 autonomous regions, and it is also classified into 7 geographical regions (North China, Northeast China, East China, South China, Central China, Southwest China and Northwest China). For each geographical region, one province (municipality or autonomous region) will be selected. Selection of the province will be in consideration of the size of MSM population in this province and the experience of local STI control staff of working with this population. In each of the seven geographical regions, one provincial capital city and one prefecture-level city will be identified as the study sites for recruiting the participants. The prefecture-level city should be at least 100 km away from the capital city and being able to recruit enough participants given the time period of the study. If there is not prefecture-level city in the selected province eligible for inclusion as a study site, the city could be selected from a neighbouring province within the geographical region. In each of the 14 study cities (figure 1), a formative assessment will be conducted to estimate the sizes of the MSM populations who most often seek sex partners at specific venues, including bars, tea houses, dance halls, public bathhouses/saunas, parks and internet sites. Each kind of venues represents a stratum of MSM to be potentially surveyed. However, considering the effect of venue typologies on risk of MSM to get STIs including HIV and the operational feasibility to recruit enough number of MSM in the study site, the venues with large size of the MSM population will be preferentially selected for the survey. However, the MSM recruited from bars, tea houses, dance halls, public bathhouses/saunas or parks should be not less than 20% and those from internet sites not more than 60% of the total sample size needed.3 We will partner with an established community-based organisation or health facility that provides a comprehensive set of services to the MSM population locally to have access to the venues.

Sample size
There are a few published studies on prevalence of CT infections among MSM and most of them were conducted in a single city in China.4–6 The prevalence varies significantly across the studies and between the anatomical sites (urethra, rectum and oropharynx) in this population. The prevalence ranges from 15% to 24% for anorectal, 3% to 5% for urethral and 4% to 6% for oropharyngeal infections. We choose the highest prevalence as it yields the highest sample size. Given an estimated prevalence of 25%, 5% and 2% for anorectal, urethral and oropharyngeal infections, respectively, and taking into account 95% confidence level and 5%, 2% and 1% precisions accordingly, the sample sizes needed are 289, 457 and 753 to estimate the prevalence in the three anatomical sites. As the prevalence of CT infections at anorectal site is the most important indicator of interest among MSM,
the sample size of at least 289 will be required for each of the seven geographical regions to do the region-specific analyses. A sum of sample sizes from the seven regions will ensure the precisions of around 1.8%, 0.9% and 0.6% for the overall estimated prevalence rates of 25%, 5% or 2% at anorectal, urethral and oropharyngeal sites, respectively. Considering about 20% non-response or drop-out rate, the sample size is increased from 289 to 347 in each region (2429 participants in total).

Participants
Inclusion and exclusion criteria
Inclusion criteria are defined as follows: men aged 16 years or above who self-reportedly engage in sexual activity with other men in the previous 6 months, are residents in the study area for more than 3 months and willing to electronically provide informed consent before filling out a questionnaire. The sexual activity refers to activity associated with penile–anal intercourse and/or oral–genital contact among men. Exclusion criteria are defined as follows: men who deny having any sexual activity with other men during the last 6 months, refuse to participate in the study by not giving informed consent, or have had participated in the current study already or administered any antibiotics in the past month. The eligible participants are requested to provide personal contact information for informing the testing results as well as providing the further follow-up. However, refusal to provide the contact information is not considered as an exclusion criterion.

Recruitment
Since there are no registers of the MSM population in each venue that can be used in this study, a convenience sampling strategy will be used for recruitment. MSM will be recruited in proportion to the estimated size of the MSM population at each stratum. For each participant, questionnaire survey and specimen collection can take place only after obtaining the informed consent.

Questionnaire interview
A questionnaire will be completed through a mobile phone app-based platform by the participants themselves. After answering a few questions to determine eligibility to participate in the survey (online supplemental appendix 1), those who are eligible to participate will be asked to provide electronic informed consent (online supplemental appendix 2) and fill out a questionnaire (online supplemental appendix 3). In the questionnaire, the participants will be invited to answer questions about their education level, occupation, marital status, sexual orientation, primary sexual role, number of sexual partners during the past 6 months and history of sexually transmitted diseases (STDs) (e.g., self-reported HIV status, prior infection with syphilis or gonorrhoea). After completing and submitting the questionnaire, an individual QR code will be generated. Then, study investigators will scan the individual QR code and the bar code on the sampling package to realise the binding of individual information, questionnaire survey and collected specimens.

Specimen collection
A total of 10–15 mL first-void urine (FVU) will be collected in a sterile urine collection jar by the participants themselves, while anorectal and pharyngeal swabs will be collected by trained staff in a private space (such as toilet, separate or temporarily building a private space with curtains) according to the standard operating procedure (online supplemental appendix 4). The collection of swabs will be conducted twice to obtain two anorectal and pharyngeal swabs. One swab of anorectal or pharyngeal swabs will be used for single-site testing and another for pooled sample testing. The sampling order was arbitrarily done for the pooled and single-site samples. Detailed instructions in Chinese on how to collect the specimens in different anatomical sites will be available for all study sites. All samples will be transported to the local designated laboratory on the same day of collection and then frozen at ~70°C. Upon completion of adequate participant recruitment, these samples will be transported along a ‘cold chain’ of refrigerators to the National Center for STD Control (NCSTD) at Nanjing.

Allowance
All participants will receive a reimbursement of 100 RMB (equivalent to US$15) for their time to complete the questionnaire survey and collect the specimens.

Laboratory assays
Preparation of pooled specimen
For preparing the pooled samples, a pharyngeal swab and a rectal swab will be inserted into the Cobas PCR Media tube, and then FVU will be pipetted into the same tube so the volume of fluid is between the recommended fill lines. The tube should be tightly recapped and inverted for five times to mix the samples. Then, the pooled samples are ready for testing or storage.

Identification of CT and NG
As in our previous study,6 the automated magnetism nucleic acid isolation method using the MagNA Pure 96 System (Roche, Switzerland) will be used to extract and purify the DNA from the pooled samples. This will be done according to the manufacturer’s instructions. The extracted DNA will be then analysed for CT and NG based on the PCR of the Cobas 4800 System using Cobas 4800 CT/NG Amplification/Detection Kit (Roche, Switzerland). The results (positive or negative amplicons) will be automatic according to the preset computer program. If the result of the pooled sample testing is positive for either CT or NG, the single-site samples from three anatomical sites (urine and swabs) will be separately tested again following the above laboratory approaches. In addition, 10% of single-site samples of participants with negative result of pooled sample testing will be tested to validate the accuracy (performance) of the pooled sample testing. The positive amplicons for CT or
NG automatically generated from the Cobas 4800 System represent a CT or NG positivity, respectively.

**Genotype of CT**

Samples yielding a positive result for CT will be sent to the NCSTD at Nanjing for ompA genotyping. As in our previous study, DNA will be extracted from the CT-positive single-site samples with the use of the DDH2O, Proteinase K and Buffer gA1 (TSINGKE, China), vortex oscillation 10 s and incubated for 15 min at 70°C. DNA will be obtained after purification and undergoing nest PCR. Nest PCR uses two pairs (instead of one pair) of PCR primers to amplify the complete fragment. For the first run, the sample DNA is used as the template, while for the second run, the template is changed to the PCR product of the first run. The primer sequence, the PCR reaction system and the amplification program will be provided by the MLST database website (http://mlstdb.bmc.uu.se) of the Uppsala University. The primers of ompA gene are shown in **table 1**. The Qingke biotechnology company (TSINGKE, China) synthesised all primers. For the first run, the CT-positive sample DNA will be used as the template. The PCR reaction system is 30 µL, including I-5× High Fidelity Master Mix (MCLAB, USA) 15 µL, outer primer F and R 1 µL, respectively; template 1 µL, filled with water to 30 µL. Amplification procedure: predegeneration at 98°C for 2 min, degeneration at 98°C for 10 s, annealing at 55°C for 10 s, stretching at 72°C for 25 s for 30 cycles, 75 repair extension for 5 min and preservation at 4°C. The PCR products will be diluted 200 times. These products will be the templates of the second run. The PCR reaction system and procedure of the second run will be the same as those of the first run. The second PCR products with gel extraction will be sequenced. The ompA sequence will be compared using BLAST on the PubMed website, and the ompA genotype (serovar) of a CT strain will be identified.

**Identification of genotypes related to LGV**

A fragment of the ompA gene (ca 1070 bp) was subsequently amplified by hemi-nested PCR using a previously reported protocol. Part of this amplicon (ca 1000 bp) was sequenced in both directions and aligned to reference sequences from GenBank representing different LGV variants: L2a (GenBank accession number: AB915594); L2 (AM884176); L2b (AM884177); L2c (EF460796); L2d (EF460797); L2e (EF460798); L2f (EU676181); L2g (EU676180); L2bV1 (JX971936); L2bV2 (KU518893); L2bV3 (KU518894); L2bV4 (KU518892); L2-L2b/Da (MN094864).

**Case management**

**Notification of testing results**

The pooled sample testing results (either positive or negative for CT or NG) will be informed to the participants in 2 weeks through the aforementioned mobile phone app-based platform if the participant has provided the personal contact information when he fills out the questionnaire. Single-site testing will be conducted only in the participants with positive pooled sample testing to identify the anatomical site-specific infections and it will take about another few days to get the results. In order to guide the participants with CT or NG, regardless of the infections in any anatomical sites, to seek for treatment timely, the pooled sample testing results will be informed to the participants.

**Treatment and partner notification**

According to the Guidelines of Clinical Management of STDs in China, the treatment for urogenital chlamydial infection should be based on results from testing for the infection. It means the treatment would not be initiated until after knowing the testing results in our study population. We will provide treatment service to the participants if they are positive for CT/NG results according to the national treatment guidelines: NG infection will be treated with ceftriaxone 1 g as a single dose while CT infection will be treated with doxycycline 100 mg orally two times per day for at least 10 days. For those with genotype of L1, L2 or L3, follow-up services to retest the specimens for cure of testing and re-treat the participant as per LGV will be conducted. Participants with positive pooled sample testing for either CT or NG will be encouraged to notify their sexual partners for CT and NG screening or presumptive treatment and participate in a prospective follow-up.

**Outcome measures**

**Infection with CT and NG**

Participants with negative results of the Cobas 4800 System for CT or NG from any anatomical site are considered to be uninfected with CT or NG. Participants with positive results of single-site sample for CT or NG are considered to be infected with CT or NG in the site, while negative

---

**Table 1 Primer pairs used for PCR amplification and sequencing of ompA**

<table>
<thead>
<tr>
<th>Primer pairs</th>
<th>Reaction system</th>
<th>Sequencing program</th>
</tr>
</thead>
<tbody>
<tr>
<td>118F Outer primer</td>
<td>5’-ATTGCTACGGACATCTTGTGC-3’</td>
<td>5’-CGGAATTGTCATTACGTGAG-3’</td>
</tr>
<tr>
<td>1163R Outer primer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ctr200F Sequencing</td>
<td>5’-TTAGGIGCTTCTTTCCAAATAYGCTCAATC-3’</td>
<td>5’-GCCAYTCATGGTARTCAATAGAGGCATC-3’</td>
</tr>
<tr>
<td>ctr254R Sequencing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOMP87 Inner primer</td>
<td>5’-TGAACCAAGCTTATGATCGAAGGA-3’</td>
<td>5’-GCAATACGGCAAGATTTTCATGATTTCATC-3’</td>
</tr>
<tr>
<td>RVS1059 Inner primer</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

results from any individual site are considered to be uninfected accordingly.

Case of LGV
An LGV case is defined in this study as a participant with laboratory confirmation to detect L genotypes associated with LGV regardless of any clinical symptoms, such as proctocolitis, associated with the disease.

Statistical analysis plan
To establish a database, one investigator will enter the data from laboratory tests into Microsoft Office Excel forms and integrate this dataset with the data from a mobile phone app-based questionnaire survey. First, statistical analyses will be carried out after the inclusion of at least 289 participants with valid questionnaire data and laboratory results. Prevalence of CT or NG will be calculated by dividing the sum of all confirmed infections by the total number of participants who participated. The prevalence will be adjusted by using the performance parameters of the pooled sample testing and a weighted prevalence will be calculated by using the proportion of recruited MSM in each geographical region relative to the estimated number of MSM in the region as the weight. Sociodemographic and behavioural data will be summarised using descriptive statistics, mean and SD, or median and IQR for continuous variables and proportion and its 95% CIs for categorical variables. Statistical tests (χ² tests or t-tests) will be used based on the type of variables. The independent variables will be the study area, age group, education level, marital status, local residency, sexual orientation, primary sexual role, number of sexual partners during the past 6 months, and self-reported HIV infection, syphilis or gonorrhoea. To identify potential predictors of the outcome of interest (infection with CT), exploratory regression model analysis (univariate followed by multivariate) will be conducted. Variables that show significant association between variables and the outcome (p≤0.1) in univariate analyses will be included in the multivariate regression model to explore the association of variables with the outcome. Using the model, adjusted OR and its 95% CIs will be estimated. P values of ≤0.05 will be considered as being statistically significant. All statistical analyses will be conducted using the IBM SPSS Statistics for Windows V.20.0 (IBM Corp) and the MedCalc for Windows V.16.8 (Mariakerke, Belgium). The study will be reported as per the Strengthening the Reporting of Observational Studies in Epidemiology guidelines.12

Patient and public involvement
Patient and public involvement (PPI) representatives worked with us to discuss the on-site implementation process of the study and refine the questions in the proposed questionnaire as well as the content of the informed consent form. However, it was hard to involve PPI representatives in other parts of the study design due to the technical complexity of an epidemiological study. In addition, PPI representatives were involved in the dissemination plans of this study.

Quality assurance, data monitoring and safety
The study coordinators and laboratory assistants in all study sites will receive 1-day training from principal investigators and experts prior to implementation of the study. The training sessions will cover possible issues raised from participants on the study, informed consent process, questionnaire survey, collection of specimens, and temporary storage of specimens at study site and their transfer to the local laboratory. To ensure comparable procedures of data collection across all participating sites, a mobile phone app-based self-administered questionnaire will be used to collect the sociodemographic and behavioural information and then the platform automatically generates an anonymous identification number to be assigned to the participant. All participating cities will update their recruitment status monthly and report them to the NCSTD to ensure that recruitment is carried out as expected. Recruitment and data collection are also monitored in regular project meetings, site visits and teleconferences.

ETHICS AND DISSEMINATION
Ethics approval
This study adheres to the principles of the Declaration of Helsinki and was reviewed and approved by the Medical Ethics Committee of the Chinese Academy of Medical Sciences Institute of Dermatology and the NCSTD in China (approval number 2021-KY-037). The study is based on voluntary participation and a written informed consent process.

Dissemination plan
The study results will be submitted for publication in peer-reviewed journals and reported in conferences. No personal information will be disclosed, and participants will not be identified when the findings of the research are published or reported publicly. The data of the study will be used as a part of the fundamental dataset for estimating the disease burden of chlamydial infections. The relevant data will be also made available for developing the national or local control programmes and used as health education materials to disseminate to the community. The dataset will be deposited in a public repository.

DISCUSSION
Currently, there are no published data that reliably estimate the national prevalence of CT infections in China. To the best of our knowledge, this is the first geographically representative study to estimate prevalence rates and to determine genotype distributions of CT infections among the MSM populations in China. Extended knowledge on the current magnitude and geography of the chlamydia epidemic among MSM in China will help
advise future national evidence-based policy development. The study also aims to investigate the specific genotypes associated with the LGV strains in this population in China. LGV outbreaks or increasing prevalence of LGV has appeared during the last decade in many countries in North America, Europe and the Pacific in the form of protococci among MSM. However, no cases with LGV-related protococci or genotypes associated with LGV have been reported from clinical practice or epidemiological studies since a case was reported by us in 2007. As reported in European countries, LGV cases are believed to be substantially underdiagnosed in China. This study design may have two major limitations on sampling strategies to be considered when the study findings are interpreted. First, the provinces or the cities within provinces to be selected for conducting the study are not randomly sampled. Selection of the province in each of seven geographical regions is only in consideration of the size of MSM population, and the study sites in the selected provinces include only the provincial capital city and one prefecture-level city where the local STI control staff have experience of working with this population. Although the MSM communities in China are mainly located in big cities, the current sampling of study sites may result in a selection bias. Second, due to operational feasibility, it is not possible to include all kinds of venues where MSM seek sexual partners in the study or recruit those MSM who may not attend any specific venues. The sampling process of study participants may further result in a selection bias. In addition, considering the operational complexity if too many kinds of specimens (urine, swabs and blood) are collected, we will not collect blood specimens for determining HIV or syphilis infection. Alternatively, we will use a questionnaire to collect self-reported HIV and syphilis status. Admittedly, the quality of this information is influenced by the participant’s recall or response bias.

In summary, this study will be useful in further understanding of the epidemiological patterns of CT and NG infections in the MSM population in China. In addition, the study will allow us to better understand the socio-demographic and behavioural predictors of the infection and the sexual network linked to transmission of the infection in this population. The findings from the study will be used as evidence-based data to contribute to the China Chlamydia Intervention Program and the STOPCT (acronym for Screening tools and strategies; Treatment regimens; Outcomes due to untreated infections; Prophylaxis; Cost-effectiveness; and Telemedicine) research plan in China.

Author affiliations
1Institute of Dermatology, Chinese Academy of Medical Sciences and Peking Union Medical College, Nanjing, Jiangsu, China
2Department of Dermatology, Affiliated Hospital of Guangdong Medical University, Zhanjiang, Guangdong, China

Contributors X-SC conceptualised the study, T-TJ, YPY and X-SC designed the study, YZ, YH, N-XC and M-QS participated in the study design. T-TJ and YPY wrote the first draft of the study protocol. X-SC made revision on the draft. YZ and YH contributed to the finalisation of the protocol. All authors reviewed and approved the final manuscript.

Funding This work was funded by research funding to X-SC from the Peking Union Medical College’s Discipline Construction Project (201910602401).

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

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

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

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

ORCID ids Ting-Ting Jiang http://orcid.org/0000-0002-0789-8816
Xiang-Sheng Chen http://orcid.org/0000-0003-3927-2065

REFERENCES
3 Huan X-ping, Yin Y-ping, Fu G-feng, et al. [Analysis on sexually transmitted diseases and the related risk factors among men who have sex with men in Jiangsu province], Zhonghua Yu Fang Yi Xue Za Zhi 2011;45:975–8.