Developing and validating a risk scoring tool for chlamydia infection among sexual health clinic attendees in Australia

A simple algorithm to identify those at high risk of Chlamydia infection

Handan Wand 1§, Rebecca Guy1, Basil Donovan1, Anna McNulty A2,3

1National Centre in HIV Epidemiology and Clinical Research, Sydney, NSW, Australia
2Sydney Sexual Health Centre, Sydney, NSW, Australia
3School of Public Health and Community Medicine, University of NSW, Kensington, Sydney, NSW, Australia

Short title: Risk scoring tool for chlamydia infection

Page count (including references and tables): 18
Word count – Text: 2717; Abstract: 250; Tables: 5

Competing Interest: none declared

§Corresponding author: Handan Wand (hwand@nchecr.unsw.edu.au)

National Centre in HIV Epidemiology and Clinical Research
Faculty of Medicine, University of New South Wales,

45 Beach Street, Coogee, NSW-2034, Sydney, Australia.
Objective: To develop and validate a risk scoring tool for chlamydia infection to identify those who are at increased risk of chlamydia infection.

Methods: We used demographic, sexual behavior information and chlamydia positivity results from more than 45,000 individuals who attended Sydney Sexual Health Centre between 1998 and 2009. Participants were randomly allocated to either the development or internal validation dataset. Using logistic regression, we created a prediction model and weighted scoring system using the development dataset and calculated the odds ratio (OR) of chlamydia positivity for participants in successively higher quintiles of score. The internal validation dataset was used to evaluate the performance characteristics of the model for five quintiles of risk scores including population attributable risk (PAR), sensitivity and specificity.

Results: In the prediction model, inconsistent condom use, increased number of sexual partners in last 3 months, genital or anal symptoms and presenting to the clinic for sexually transmitted infections (STIs) screening or contact of an STI case were consistently associated with increased risk of chlamydia positivity in all groups. High scores (upper quintile) were significantly associated with increased risk of chlamydia infection. A cut-point score of 20 or higher distinguished a ‘increased risk’ group with a sensitivity of 95%, 67% and 79% among heterosexual men, women and MSM, respectively.

Conclusion: The scoring tool may be included as part of a health promotion and/or clinic website to prompt those who are at increased risk of chlamydia infection which may potentially lead to increased uptake and frequency of testing.

Key Words: Chlamydia infection, risk prediction
INTRODUCTION

Chlamydia infection is highly prevalent in young heterosexuals and men who have sex with men (MSM) in Australia with prevalence estimates of 4-6% in both populations [1,2]. The majority of chlamydia infections are asymptomatic. Chlamydia is associated with sequelae such as pelvic inflammatory disease (PID) and infertility in women and proctitis in men [3-6]. Also in MSM, chlamydia re-infection of the rectum has been associated with an increased risk of HIV seroconversion [7].

The number of chlamydia notifications continues to increase steadily each year among MSM and young heterosexual men and young women in Australia [8,9] as in many other countries. A major public health challenge is therefore to identify individuals at risk of chlamydia and facilitate testing and treatment before the development of chlamydia sequelae and onward transmission to others. Clinical guidelines in Australia recommend annual chlamydia testing in <25 year olds, annual HIV and STI testing for all MSM and 3-6 monthly testing for high-risk MSM reporting more than ten sexual partners in the last six months, unprotected sex and other specific risk behaviours [10,11].

Clinical risk prediction approaches that can capture a continuous risk spectrum have been used in decision making about public health and clinical care and have been proposed as an alternative to diagnosis for some diseases in various contexts [12,13]. Our study aimed to develop and validate a simple scoring tool to assess the risk of chlamydia infection using demographic and sexual risk behavior information collected from over 45,000 individuals who attended Sydney Sexual Health Centre (SSHC) between 1998 and 2009.

METHODS
Study population

The study population consists of 45,902 men and women who visited SSHC during the period of 1998-2009. A standard medical record form was used to collect demographic and sexual behavior information from all new attendees and a sexual health screen undertaken. Since 1998 SSHC has actively triaged those at higher risk of STIs into the service. SSHC also targets sex workers from culturally and linguistically diverse (CALD) backgrounds through interpreter facilitated sex worker clinics.

For this analysis, the demographic and sexual behavior information were extracted from the medical records system including the anonymous patient identifier, age, gender, postcode, country of birth, arrival in Australia (if overseas-born), marital status, alcohol use, condom use, number of male/female sex partners in the last 3 and 12 months, sex overseas in the past 12 months, reason for attendance, self-reported past chlamydia diagnoses, perceived HIV status and the current HIV/STI test results.

Statistical Analyses

A split-sample method was used to develop a risk equation and scoring system with internal validation for each study population. Participants were randomly allocated to either the development (~67%) or internal validation (~33%) sample datasets within each group.

Development dataset

Logistic regression was used to create a predictive model based on the development dataset which included 11,354, 6,800 and 12,700 MSM, heterosexual men and women, respectively. We evaluated a range of socio-demographic and sexual behavior variables as potential determinants of chlamydia infection including age, country of birth (Australia vs. other countries), language spoken at home (English vs. others), marital status [married/defacto vs. others (divorced/widow/unknown)], culturally and linguistically diverse (CALD) (not born in Australia and not speak English at home), travelers (not born in Australia and been Australia for less than 2 years or those who identify themselves as ‘traveler’), area of residence, alcohol use, number of sexual partners in the past 3 months, condom use in the past 3 months, sex overseas in the past year, current sex work, reason for presentation, anal/genital symptom, past chlamydia
diagnoses, perceived HIV status. All analyses were stratified by sexual identities: MSM, heterosexual men and women.

Ethical approval for the study was obtained from the South Eastern Sydney and Illawarra Area Health Service Human Research Ethics Committee.

We used descriptive statistics to characterize the groups according to chlamydia status: mean and standard deviation (SD) for continuous variables and percentages for categorical variables. Logistic regression was used to create a predictive model based on the development data set. We used all non-missing observations available in the relevant analyses as only a small proportion of observations had any missing data. All analyses were conducted using SAS statistical software version 9.2 (SAS Institute Inc. Cary, North Carolina) and STATA 10.0 (College Station, Texas).

**Derivation of a screening score**

Using the development data sets for the three subgroups, we investigated a comprehensive list of predictors known to be potentially associated with chlamydia infection in an initial model. Specifically, we included the main effects of all variables listed in Table 1. We first analyzed the univariate associations between each variable and being diagnosed with STIs in each subgroup separately. Backward elimination was used to reach the final multivariate model, in which factors with the largest $P$ value were sequentially deleted until only significant predictors remained. We then created a weighted scoring system by rounding all regression coefficients up to the nearest integer (that is, the smallest integer greater than the estimate). This method was based on the $\beta$-coefficients (or log of the odds ratios) rather than odds ratios, which can be excessively influenced by only a few factors [26]. Once the final model was defined, we created integer weights for each variable. We calculated these weights by multiplying the model coefficients by 10. Using the rounded weights in the risk function, we estimated the participant-specific probabilities of chlamydia positivity and characterized the degrees of risk based on cutoff points of the probability distribution.

**Cross-sectional internal validation**

The prediction model was evaluated in the three cross-sectional internal validation datasets of 3,805 MSM, 5,313 heterosexual men and 7,084 women. We conducted various analyses to check the sensitivity and robustness of the new screening score. We computed standard validation
measures: the proportion of those tested positive for chlamydia infection, sensitivity, specificity, positive likelihood and negative likelihood ratio and the area under the receiver-operating characteristic curve (AUC) [8] as discrimination statistics. We also assessed the diagnostic characteristics of different cut-points based on the total score in the development as well as validation datasets. The purpose of this analysis was to assess whether the combination of risk factors under consideration could predict those at increased risk with acceptable accuracy.

Population attributable risk

We then estimated the PAR, which estimates the percentage of chlamydia infections that would not have occurred if all the participants had been in the “lowest risk” (first quintile) category of the risk score. We calculated population attributable risks by using previously described method [9] that were elaborated for this study design and are appropriate for use with multivariate adjusted relative risks.

RESULTS

Table 1 summarizes participant characteristics by each group. The overall prevalence of chlamydia was 6%, 7%, and 5% for MSM, heterosexual men and women respectively. MSM were more likely to be Australian born and live in metropolitan Sydney. More than 30% of the females were from culturally and linguistically diverse background (CALD) backgrounds compared to 13% of heterosexual men and MSM. Approximately 50% of the females were also classified as “travelers” compared to 38% and 27% heterosexual men and MSM respectively. Although excess alcohol intake and current smokers were more common among heterosexual men and women compared to MSM, more MSM reported ever injecting drug use. Approximately 50% of women reported being in full time employment and 20% of them identified as being a sex worker. More heterosexual men reported that they had had sex in Asia in last 12 months. Inconsistent condom use in last 3 months and presenting with genital or anal symptoms were more common among heterosexual men and women compared to MSM. The primary reason for making an appointment was testing for sexually transmitted infections in all groups, however, presentation for HIV testing was more common among MSM compared to heterosexual men and women. Consistent with this, approximately 50% of heterosexual men and women also did not know their HIV status compared to 22% of the MSM.
**Prediction model**

Table 2 presents the final multivariate logistic regression model derived from the development dataset for each group. Independent predictors of chlamydia infection in MSM were younger age, inconsistent condom use, increased number of male sexual partners in the past 3 months, anal/genital symptoms and presenting for STI screening or a contact of an STI case.

Independent predictors of chlamydia infection in heterosexual men were being single, CALD background, being unsure about HIV status, inconsistent condom use, increased number of female sexual partners in the past 3 months, anal/genital symptoms and presenting for STI screening or a contact of an STI case.

Independent predictors of chlamydia infection in women were being single, CALD background, being unsure about HIV status, inconsistent condom use, anal/genital symptoms, presenting for STI screening or contact of an STI case.

**Internal validation**

The variables age and number of male/female sexual partners required multiple categories to capture the risk gradient, whereas other risk factors were binary. The risk factors collectively yielded an AUC 0.71 (95% CI: 0.69, 0.73) for MSM, 0.74 (95% confidence interval (CI): 0.72, 0.75) for heterosexual men and 0.72 (95% CI: 0.70, 0.74) for women. There were no statistically significant interactions detected between the sexual risk factors and the age groups (data not shown).

Table 3 shows the ORs from the logistic regression models for the quintiles of the risk scores in development and validation datasets. The ORs (95% CI) of chlamydia positivity for participants in successively higher quintiles of STI score were 1.79 (1.23, 2.60), 2.96 (2.10, 4.15), 4.56 (3.30, 6.30) and 8.80 (6.43, 12.02) (MSM); 2.53 (1.76,3.63), 4.21 (2.97,5.98), 6.82 (4.84, 9.60) and 14.17 (10.20, 19.68) (heterosexual men); 2.50 (1.67, 3.76), 3.70 (2.51, 5.43), 4.59 (3.11, 6.78) and 12.33 (8.55,17.78) (heterosexual women). There was a linear trend towards increasing chlamydia positivity with increasing score regardless of the groups for development and validation datasets (trend, p-value<0.001, all).
We also estimated PARs (95% CI) for the upper four quintiles of the scores. Results showed that 73% (69%, 76%) of infections in MSM, 80% (77%, 82%) of infections in heterosexual men and 78% (74%, 81%) of infections in women would be avoided if the participants who were in the upper four quintiles of the STI scores were in the lowest quintile. Results from the validation dataset were consistent with the results from the development dataset.

We performed additional analyses to calculate probability of being infected by quintiles of the risk score for each group separately and presented in Table 4. The chance of chlamydia positivity increased in the higher quintiles of the score. In validation data set, MSM, heterosexual men and women who were in the highest quintile of the risk score were estimated to have at least a 24%, 35% and 42% chance of being infected. An increased probability of chlamydia positivity was clearly associated with increasing scores regardless of the group under consideration.

We also assessed the diagnostic characteristics of cut-points (according to first, second, third and the fourth quintiles in overall population) for total score in the development as well as the validation datasets (Table 5). An increased risk score was associated with increasing sensitivity for chlamydia infection. A cut-point score of 20 or higher distinguished a ‘higher risk’ group with a sensitivity of 95%, 67% and 79% among heterosexual men, women and MSM respectively.

DISCUSSION

In this study, we have developed a chlamydia risk scoring tool based on data from more than 45,000 men and women who attended SSHC during the period of 1998-2009. The tool was validated to accurately identify those at increased risk of chlamydia infection. Our methodology made use of a range of coexisting risk factors that were identified by a rigorous statistical approach in order to accurately identify the most relevant risk factors for chlamydia infection.

Developing a risk assessment tools that identifies, quantifies and characterizes risks may lead to improved knowledge about chlamydia and increased testing for STIs. This is particularly relevant because many infections are asymptomatic and individuals may be unaware that are at risk and/or have the infection. For example, our current study found a higher percentages of heterosexual males and females were unsure of their HIV status compared to MSM (47%, 48% and 22% for heterosexual men, women and MSM respectively) and those who were not aware of
their HIV status were determined to be at high risk for chlamydia infection (OR: 1.38, p-value=0.001 and OR: 1.54, p-value<0.001 for heterosexual men and women respectively).

This study has several strengths. It is the first study to utilize statistical methods to derive a sensitive and locally-specific risk assessment tool to identify, quantify and characterize risks of various groups in Australia. Risk assessment methods or prediction models ideally should be derived from large representative samples. Our study used 13 years of data from more than 45,000 men and women to develop the risk assessment tool suggested in this study. Our risk calculation was based on a statistical method that yielded a systematic scoring system for carefully selected predictors, guided not only by scientific evidence but also feasibility perspectives. However, our study is limited by its retrospective nature and the self reported measures of the sexual risk factors which may be subject to measurement error/misclassification. The study population was based on clinic attendees who are triaged into the service based on risk assessment and/or presence of symptoms as demonstrated by the positivity of 6%, 7%, and 5% for MSM, heterosexual men and women respectively compared to chlamydia prevalence estimates of 3% to 5% in community based studies [12]. Finally it is possible that chlamydia infection might have acquired prior to the sexual risk behavior that preceded the clinic visit as chlamydia infection can persist for an average of 12 months if untreated [13].

We envisage that the chlamydia risk scoring tool developed in this study will be adapted for interactive community and/or clinic websites and the interface and website will be designed and calibrated for use by relevant populations including people from CALD backgrounds who were at higher risk for chlamydia infection. The screening tool will also be piloted in primary care clinics setting targeting those at higher risk for infection(s).

<table>
<thead>
<tr>
<th>Key Messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>The authors created a risk assessment tool that allows people to estimate their own chlamydia risk score based on simple non-invasive variables.</td>
</tr>
<tr>
<td>The tool described here will potentially provide a simple and cost-effective method of identifying and alerting individuals who would benefit from</td>
</tr>
</tbody>
</table>
This tool may be included as part of a health promotion and/or clinic website.

This tool may potentially lead to increased uptake and frequency of testing.

In conclusion, we believe the screening tool described here will provide a simple and cost-effective method of identifying and alerting individuals who would benefit from chlamydia screening with notable predictive validity. Self identification, if widely practiced, could be an effective method of case ascertainment and may encourage uptake of screening.
REFERENCES


[13] Geisler W. Duration of untreated, uncomplicated Chlamydia trachomatis genital infection and factors associated with chlamydia resolution: a review of human studies JUID 2010:201; S104-S113
Developing and validating a risk scoring tool for chlamydia infection among sexual health clinic attendees in Australia: a simple algorithm to identify those at high risk of chlamydia infection

Handan Wand, Rebecca Guy, Basil Donovan and Anna McNulty

*BMJ Open* 2011 1: originally published online February 23, 2011
doi: 10.1136/bmjopen-2010-000005

Updated information and services can be found at:
http://bmjopen.bmj.com/content/1/1/e000005

These include:

**Supplementary Material**
Supplementary material can be found at:
http://bmjopen.bmj.com/content/suppl/2011/02/21/bmjopen-2010-000005.DC1

**References**
This article cites 12 articles, 3 of which you can access for free at:
http://bmjopen.bmj.com/content/1/1/e000005#BIBL

**Open Access**
This is an open-access article distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the license. See: http://creativecommons.org/licenses/by-nc/2.0/ and http://creativecommons.org/licenses/by-nc/2.0/legalcode.

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Topic Collections**
Articles on similar topics can be found in the following collections

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/
Notes