BMJ open accessible medical research

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 Guy,[1] Basil Donovan,[1] Anna McNulty[2,3]

To cite: Wand H, Guy R, Donovan B, *et al*. 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. *BMJ Open* 2011;**1**:e000005. doi:10.1136/bmjopen-2010-000005

► Prepublication history for this paper is available online. To view these files please visit the journal online (http://bmjopen.bmj.com).

Received 26 September 2010
Accepted 6 January 2011

[1]National Centre in HIV Epidemiology and Clinical Research, Sydney, New South Wales, Australia
[2]Sydney Sexual Health Centre, Sydney, New South Wales, Australia
[3]National Centre in HIV Social Research, University of New South Wales, Sydney, New South Wales, Australia

Correspondence to
Dr Handan Wand;
hwand@nchecr.unsw.edu.au

ABSTRACT

Objective: To develop and validate a risk scoring tool to identify those who are at increased risk of chlamydia infection.

Methods: We used demographic data, sexual behaviour 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 data set. Using logistic regression, we created a prediction model and weighted scoring system using the development data set and calculated the odds ratio of chlamydia positivity for participants in successively higher quintiles of score. The internal validation data set was used to evaluate the performance characteristics of the model for five quintiles of risk scores including population attributable risk, 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 screening or being a contact of a sexually transmitted infection case were consistently associated with increased risk of chlamydia positivity in all groups. High scores (upper quintiles) 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 men who have sex with men (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.

ARTICLE SUMMARY

Article focus

- The authors created a risk assessment tool that allows people to estimate their own chlamydia risk score based on simple non-invasive variables.

Key messages

- The tool described here will potentially provide a simple and cost-effective method of identifying and alerting individuals who would benefit from chlamydia screening.
- 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.

Strengths and Limitations

- This is the first study to utilize statistical methods to derive a locally-specific assessment tool using 12 years of data from more than 45 000 men and women.
- The Study population was sexual health clinic attendees who are likely to be at higher risk for Chlamydia infection compared to the general population.

INTRODUCTION

Chlamydia infection is highly prevalent in young heterosexuals and men who have sex with men (MSM) in Australia, with prevalence estimates of 3–5% in both populations.[1 2] The majority of chlamydia infections are asymptomatic. Chlamydia is associated with sequelae such as pelvic inflammatory disease 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 sexually transmissible infection (STI) testing for all MSM and 3–6 monthly testing for high-risk MSM reporting more than 10 sexual partners in the last 6 months, unprotected sex and other specific risk behaviours.[8]

Clinical risk prediction approaches that can capture a continuous risk spectrum have been used in public health and clinical care decision making and have been proposed as an alternative to diagnosis for some diseases in various contexts.[11 12] Our study aimed to develop and validate a simple scoring tool to assess the risk of chlamydia infection using demographic and sexual risk behaviour 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 1998–2009. A standard medical record form was used to collect demographic and sexual behaviour information from all new attendees and a sexual health screen was 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 behaviour information was extracted from the medical records system including the anonymous patient identifier, age, gender, postcode, country of birth, date of arrival in Australia (if born overseas), 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 data sets within each group.

Development data set

Logistic regression was used to create a predictive model based on the development data set which included 11 354, 6800 and 12 700 MSM, heterosexual men and women, respectively. We evaluated a range of socio-demographic and sexual behaviour 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/widowed/unknown)), CALD (not born in Australia and not speaking English at home), travellers (not born in Australia and in Australia for less than 2 years or those who identify themselves as 'travellers'), 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 symptoms, past chlamydia diagnoses and perceived HIV status. All analyses were stratified by sexual identity (MSM, heterosexual man or woman).

We used descriptive statistics to characterise the groups according to chlamydia status: mean and 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 v 9.2 (SAS Institute) and STATA 10.0.

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 analysed the univariate associations between each variable and being diagnosed with STIs in each sub-group 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 (ie, the smallest integer greater than the estimate). This method was based on the β coefficients (or log of the ORs) rather than ORs, which can be excessively influenced by only a few factors.[11] 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 characterised the degrees of risk based on cut-off points of the probability distribution.

Cross-sectional internal validation

The prediction model was evaluated in the three cross-sectional internal validation data sets of 3805 MSM, 5313 heterosexual men and 7084 women. We conducted various analyses to check the sensitivity and robustness of the new screening score. We computed standard validation measures for 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)[13] as discrimination statistics. Akaike information criteria were evaluated as model fit statistics. The Hosmer–Lemeshow goodness-of-fit test was also performed. We also assessed the diagnostic characteristics of different cut-

Table 1 Characteristics of the population

Characteristics	MSM, N = 10 154	Heterosexual male, N = 16 667	Heterosexual female, N = 19 081
Age at visit, n (%)			
<25 years	2179 (21)	4575 (27)	7516 (39)
25–29 years	2649 (26)	4851 (29)	5905 (31)
30–39 years	3342 (33)	4452 (27)	4175 (22)
40+ years	1984 (20)	2789 (17)	1485 (8)
Never been married	8408 (84)	12 247 (75)	14 520 (77)
Country of birth, n (%)			
Australia	4986 (49)	6459 (39)	5735 (30)
England	1102 (11)	3108 (19)	2561 (13)
Asian countries*	498 (5)	339 (2)	1277 (7)
Missing	3568 (35)	6761 (41)	9508 (50)
Usually live in, n (%)			
Metropolitan Sydney/eastern suburbs	6613 (65)	10 258 (62)	10 417 (57)
English spoken at home, n (%)	8781 (86)	14 387 (86)	12 922 (68)
Travellers†, n (%)	2775 (27)	6265 (38)	9766 (51)
CALD‡, n (%)	1276 (13)	2155 (13)	6003 (31)
Employment status, n (%)			
Employed	6737 (67)	10 604 (65)	9173 (49)
Unemployed	1021 (10)	2210 (14)	2327 (12)
Student	1555 (16)	2118 (13)	4665 (25)
Current sex worker, n (%)	91 (1)	41 (<1)	3745 (20)
Current smoker, n (%)	2969 (29)	5096 (31)	6627 (35)
Excess alcohol§, n (%)	429 (4)	1453 (9)	2526 (13)
Injecting drug use ever, n (%)	723 (7)	630 (4)	595 (3)
Sex overseas (Asia) (last 12 months), n (%)	895 (9)	2320 (14)	1920 (10)
Inconsistent condom use (last 3 months), n (%)	3987 (40)	3389 (21)	3579 (19)
Male sex partners (last 3 months), n (%)			
None	1193 (12)	–	2855 (15)
One	2828 (28)	–	11 610 (61)
Two	1679 (17)	–	3013 (16)
Three or more	7860 (78)	–	1498 (7)
Female sex partners (last 3 months), n (%)			
None	–	1989 (12)	–
One	–	7642 (46)	–
Two or more	–	6913 (41)	–
Reason for presentation, n (%)			
HIV testing	1194 (12)	1154 (7)	843 (4)
STI test	3515 (35)	4131 (25)	6282 (33)
STI contact	568 (6)	796 (5)	761 (4)
Prophylaxis	305 (3)	29 (<1)	17 (<1)
Genital/anal symptoms, n (%)	3392 (33)	8998 (54)	7891 (41)
Past diagnosis, n (%)			
Chlamydia	1450 (14)	2300 (14)	1796 (9)
Gonorrhoea	1587 (16)	659 (4)	264 (1)
Syphilis	365 (4)	84 (1)	71 (<1)
HPV	975 (10)	1777 (11)	1882 (10)
Genital herpes	481 (5)	629 (4)	1117 (6)
Not tested/not sure of HIV status, n (%)	2199 (22)	7894 (47)	9168 (48)
Current diagnosis, n (%)			
Chlamydia	656 (6)	1124 (7)	973 (5)
Gonorrhoea	516 (5)	207 (1)	79 (<1)
Syphilis	109 (1)	10 (1)	12 (<1)
HPV	628 (6)	1623 (10)	1172 (6)
Genital herpes	183 (2)	594 (4)	749 (4)
HIV	113 (1)	11 (<1)	17 (<1)

*China (excludes SARs and Taiwan), Indonesia, South Korea, Malaysia, Philippines, Singapore and Thailand.
†Born outside Australia and arrived in Australia in last 2 years.
‡Born outside Australia and does not speak English at home.
§Average alcohol intake of 280 g for men and 140 g for women per week.
CALD, culturally and linguistically diverse; HPV, human papillomavirus; MSM, men who have sex with men; STI, sexually transmissible infection.

Table 2 Multivariate logistic regression: ORs and 95% CIs for chlamydia infection

	MSM			Heterosexual males			Heterosexual females		
Development data set	OR (p value)	β Coefficient (×10)	Score	OR (p value)	β Coefficient (×10)	Score	OR (p value)	β Coefficient (×10)	Score
Age (years)									
<25	1.47 (0.005)	3.9	4	–			–		
25–29	1.58 (<0.001)	4.6	5	–					
30–39	1.42 (0.006)	3.5	4	–					
40+	1	–	0	–					
Marital status									
Married/defacto	–			1	1		1		
Never married/divorced/other	–			1.48 (<0.001)	4.0	4	1.25 (0.05)	2.0	2
Usually live in									
Metropolitan Sydney/eastern suburbs	1	–	0	–			1	–	
Interstate/another part of New South Wales	1.30 (0.003)	2.6	3				1.33 (<0.001)	2.9	3
CALD*				1.36 (0.007)	3.0	3	1.43 (<0.001)	3.6	4
Travellers (backpackers)†				1.43 (<0.001)	3.6	4	–		
Sex in Asia				1.66 (<0.001)	5.1	5	–		
Current sex worker (vs never/past)	–			–			1.73 (<0.001)	5.5	6
Current smoker (vs never/past)	1.19 (0.05)	1.8	2	1.30 (0.001)	2.6	3	–		
Symptoms									
No symptoms	1	–	0	1			1	0	
Genital or anal	2.86 (<0.001)	10.5	11	2.54 (<0.001)	9.3	9	1.45 (<0.001)	3.7	4
Reason for presentation									
Other reasons	1	–	0	1			1	0	0
STI test	1.30 (0.011)	2.6	3	1.35 (0.029)	3.0	3	1.26 (0.029)	2.3	2
STI contact	5.46 (<0.001)	17.0	17	7.56 (<0.001)	20.0	20	9.23 (<0.001)	22.2	22
Contraception use (females only)									
Others‡	–			–			1	0	
Oral contraception							1.50 (<0.001)	4.0	4
No contraception							1.58 (0.009)	4.6	5
Inconsistent condom use (in last 3 months)‡	1.51 (<0.001)	4.1	4	2.11 (<0.001)	7.5	8	1.51 (<0.001)	4.1	4
Past chlamydia	–			1.51 (<0.001)	4.1	4	–		
Number of male sex partners (in last 3 months)							–		
None	1	–		–			1	0	0
One	1.63 (0.008)	4.9	5	–			1.31 (0.09)	2.7	3
Two	2.58 (<0.001)	9.5	10	–			1.74 (0.02)	5.5	6
Three or more	3.13 (<0.001)	11.6	12				2.52 (<0.001)	9.2	9

Continued

Table 2 Continued

Development data set	MSM			Heterosexual males			Heterosexual females		
	OR (p value)	β Coefficient (×10)	Score	OR (p value)	β Coefficient (×10)	Score	OR (p value)	β Coefficient (×10)	Score
Number of female sex partners (in last 3 months)									
None/one	—			1			—		
Two or more	—			2.37 (<0.001)	8.6	9	—		
Knowledge of HIV status									
Others (tested at least once)	—			1			1		
Not tested/unsure				1.38 (0.001)	3.2	3	1.54 (<0.001)	4.4	4

*Born outside Australia and English not spoken at home.
†Born outside Australia and arrived in Australia in last 2 years.
‡Never, sometime, usually, unknown.
CALD, culturally and linguistically diverse; MSM, men who have sex with men; STI, sexually transmissible infection.

points based on the total score in the development as well as the validation data sets. 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 population attributable risk (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 PAR by using previously described methods[14] that were elaborated for this study design and are appropriate for use with multivariate adjusted relative risks.

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

RESULTS

Table 1 summarises participant characteristics by 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 CALD backgrounds compared to 13% of heterosexual men and MSM. Approximately 50% of females were also classified as travellers compared to 38% and 27% of 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 the last 12 months. Inconsistent condom use in the 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 STI 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 MSM.

Prediction model

Table 2 presents the final multivariate logistic regression model derived from the development data set 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 being 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 being a contact of an STI case. The Hosmer–Lemeshow goodness-of-fit test showed no lack of fit for the three fitted models (p>0.21 in all models).

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 being a 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 of 0.71 (95% CI 0.69 to 0.73) for MSM, 0.74 (95% CI 0.72 to 0.75) for heterosexual men and 0.72 (95% CI 0.70 to 0.74) for women. No statistically significant interactions were detected between the sexual risk factors and the age groups.

Table 3 shows the odds ratios from the logistic regression models for the quintiles of the risk scores in the development and validation data sets. The ORs (95% CI) of chlamydia positivity for participants in successively higher quintiles of STI score were: 1.79 (1.23 to 2.60), 2.96 (2.10 to 4.15), 4.56 (3.30 to 6.30) and 8.80 (6.43 to 12.02) for MSM; 2.53 (1.76 to 3.63), 4.21 (2.97 to 5.98), 6.82 (4.84 to 9.60) and 14.17 (10.20 to 19.68) for heterosexual men; and 2.50 (1.67 to 3.76), 3.70 (2.51 to 5.43), 4.59 (3.11 to 6.78) and 12.33 (8.55 to 17.78) for heterosexual women. There was a linear trend towards increasing chlamydia positivity with increasing score regardless of group for the development and validation data sets (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% to 76%) of infections in MSM, 80% (77% to 82%) of infections in heterosexual men and 78% (74% to 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 data set were consistent with results from the development data set.

We performed additional analyses to assess the diagnostic characteristics of various cut-points of the total score in the overall study population (table 4). For example, among heterosexual men, the predictive value of the screening criteria for a cut-point score of 20 or higher was approximately 10%. Although it is crucial to determine the best cut-point to alert those at highest risk for infection, a cut-point of ≥20 or higher demonstrated excellent sensitivity among MSM and heterosexual males (80.0% and 96.8%, respectively) and acceptable sensitivity among heterosexual women (70.0%).

DISCUSSION

In this study, we have developed a chlamydia risk scoring tool based on data from more than 45 000 men and

Table 3 ORs and 95% CIs for being diagnosed with chlamydia infection by quintiles of chlamydia risk scoring

	MSM		Heterosexual males		Heterosexual females	
	OR (95% CI)	p Value*	OR (95% CI)	p Value*	OR (95% CI)	p Value*
Development data set						
Chlamydia risk score†						
1st Quintile	1	<0.001	1	<0.001	1	<0.001
2nd Quintile	1.79 (1.23 to 2.60)		2.53 (1.76 to 3.63)		2.50 (1.67 to 3.76)	
3rd Quintile	2.96 (2.10 to 4.15)		4.21 (2.97 to 5.98)		3.70 (2.51 to 5.43)	
4th Quintile	4.56 (3.30 to 6.30)		6.82 (4.84 to 9.60)		4.59 (3.11 to 6.78)	
5th Quintile	8.80 (6.43 to 12.02)		14.17 (10.20 to 19.68)		12.33 (8.55 to 17.78)	
AUC (95% CI)	0.75 (0.72 to 0.76)		0.72 (0.70 to 0.73)		0.73 (0.70 to 0.74)	
Population attributable risk for chlamydia risk score (%)‡ (95% CI)						
Lowest vs upper four quintiles	73% (0.69 to 0.76)		80% (0.77 to 0.82)		78% (0.74 to 0.81)	
Validation data set						
Chlamydia risk score†						
1st Quintile	1	<0.001	1	<0.001	1	<0.001
2nd Quintile	1.80 (1.17 to 2.75)		2.40 (1.30 to 4.42)		1.62 (1.01 to 2.59)	
3rd Quintile	2.29 (1.54 to 3.40)		4.67 (2.64 to 8.27)		2.80 (1.82 to 4.29)	
4th Quintile	3.93 (2.70 to 5.73)		6.81 (3.88 to 11.95)		4.07 (2.66 to 6.24)	
5th Quintile	6.57 (4.52 to 9.54)		12.71 (7.37 to 21.90)		8.10 (5.34 to 12.20)	
AUC (95% CI)	0.74 (0.72 to 0.75)		0.71 (0.69 to 0.73)		0.72 (0.70 to 0.74)	
Population attributable risk for chlamydia risk score (%) (95% CI)						
Lowest vs upper four quintiles	67% (0.62 to 0.72)		84% (0.81 to 0.87)		68% (0.63 to 0.74)	

*Test for trend.
†Chlamydia risk score was as follows: for heterosexual males (median: 26) <20 for 1st quintile, 20–25 for 2nd quintile, 26–30 for 3rd quintile, 31–35 for 4th quintile and 36+ for 5th quintile; for MSM (median: 20) <15 for 1st quintile, 15–18 for 2nd quintile, 19–22 for 3rd quintile, 23–27 for 4th quintile and 28+ for 5th quintile; and for females (median: 27) <13 for 1st quintile, 13–16 for 2nd quintile, 17–20 for 3rd quintile, 21–24 for 4th quintile and 25+ for 5th quintile.
AUC, area under the curve; MSM, men who have sex with men.

Table 4 Selection criteria for screening for chlamydia infection

Score cut-points*	MSM				Heterosexual males				Heterosexual females			
	Sensitivity†	Specificity‡	Fraction positive§	PPV¶	Sensitivity†	Specificity‡	Fraction positive§	PPV¶	Sensitivity†	Specificity‡	Fraction positive§	PPV¶
≥5	99.4%	2.5%	99.0%	11.1%	100.0%	0.0%	100.0%	7.6%	100.0%	1.8%	99.8%	5.9%
≥10	98.0%	9.0%	98.0%	11.7%	100.0%	0.5%	99.3%	8.6%	98.4%	11.8%	97.5%	6.4%
≥15	93.0%	24.0%	93.0%	13.0%	99.3%	11.2%	99.0%	9.2%	90.3%	33.3%	88.5%	7.6%
≥20	80.0%	52.0%	77.0%	18.0%	96.8%	23.3%	96.0%	10.1%	70.0%	62.3%	67.0%	9.8%
≥25	57.0%	75.0%	54.0%	23.3%	88.8%	42.1%	87.0%	12.0%	43.1%	86.1%	40.0%	15.2%
≥30	33.4%	88.5%	31.0%	27.1%	74.4%	62.2%	72.0%	14.6%	27.3%	95.2%	25.0%	24.5%
≥35	12.0%	97.9%	10.0%	40.1	50.8%	82.5%	48.0%	20.0%	19.4%	97.9%	17.0%	33.7%
≥40	3.4%	99.7%	3.0%	56.7%	29.0%	93.2%	26.0%	23.7%	11.8%	99.2%	10.0%	42.9%

*For MSM: 4×(age<25)+5×(age 25–29)+4×(age 30–39)+3×residence (not metropolitan/interstate)+0×(never married/divorced/other)+0×(CALD)+0×(travellers)+0×(sex in Asia)+0×(current sex worker)+2×(current smoker)+11×(genital or anal symptoms)+3×(STI testing)+17×(STI contact)+4×(inconsistent condom use)+0×(past chlamydia)+5×(one male sex partner in last 3 months)+10×(two male sex partners in last 3 months)+12×(three or more male sex partners in last 3 months)+0×(two or more female sex partners in last 3 months)+0×(knowledge of HIV status: not tested/unsure).
For heterosexual males: 0×(age<25)+0×(age 25–29)+0×(age 30–39)+4×(never married/divorced/other)+0×residence (not metropolitan/interstate)+3×(CALD)+3×(travellers)+5×(sex in Asia)+0×(current sex worker)+3×(current smoker)+9×(genital or anal symptoms)+3×(STI testing)+20×(STI contact)+8×(inconsistent condom use)+4×(past chlamydia)+9×(two or more female sex partners in last 3 months)+3×(knowledge of HIV status: not tested/unsure).
For females: 0×(age<25)+0×(age 25–29)+0×(age 30–39)+2×(never married/divorced/other)+3×residence (not metropolitan/interstate)+4×(CALD)+0×(travellers)+0×(sex in Asia)+6×(current sex worker)+2×(current smoker)+4×(genital or anal symptoms)+2×(STI testing)+22×(STI contact)+4×(contraception-oral)+5×(contraception-none)+4×(inconsistent condom use)+0×(past chlamydia)+3×(one male sex partner in last 3 months)+6×(two male sex partners in last 3 months)+9×(three or more male sex partners in last 3 months)+0×(two or more female sex partners in last 3 months)+4×(knowledge of HIV status: not tested/unsure).
†Percentage of detected chlamydia infection in the population (heterosexual males, MSM or heterosexual females).
‡Percentage of chlamydia negative participants in the population (heterosexual males, MSM or heterosexual females) who would not be screened justly.
§Percentage of the total population (heterosexual males, MSM or heterosexual females) eligible for screening under the given selection criterion.
¶Prevalence of chlamydia infection in the screened population (heterosexual males, MSM or heterosexual females).
CALD, culturally and linguistically diverse; MSM, men who have sex with men; PPV, positive predictive value; STI, sexually transmissible infection.

women who attended SSHC during the period 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 determine the most relevant risk factors for chlamydia infection.

Developing a risk assessment tool that identifies, quantifies and characterises 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 they are at risk and/or have the infection. For example, our current study found 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=0.001 and OR 1.54, p<0.001 for heterosexual men and women, respectively).

This study has several strengths. It is the first study to utilise statistical methods to derive a locally-specific risk assessment tool to identify, quantify and characterise the risks of various groups in Australia with acceptable sensitivity. Risk assessment methods or prediction models ideally should be derived from large representative samples. Our study used 12 years of data from more than 45 000 men and women to develop the suggested risk assessment tool. Our risk calculation was based on a statistical method that yielded a systematic scoring system for carefully selected predictors, guided by both scientific evidence and feasibility perspectives. However, our study is limited by its retrospective nature and the self-reported measures of the sexual risk factors and anal/genital symptoms 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 the presence of symptoms as demonstrated by the positivity of 6%, 7% and 5% for MSM, heterosexual men and women, respectively. When we restricted the analyses to those younger than 20 years of age, chlamydia positivity rates were estimated to be 11%, 8% and 7% for MSM, heterosexual male and females, respectively, compared to 3%–5% among young MSM, heterosexual men and women in community-based studies.[8] It is also possible that chlamydia infection might have been acquired prior to the sexual risk behaviour that preceded the clinic visit as chlamydia infection can persist for an average of 12 months if untreated.[9] Finally, risk prediction models apply primarily to groups defined by a set of clinically relevant variables rather than directly to individuals. This is a limitation common to all risk prediction models.[10] Indeed, prevention of chlamydia infection may require population-based interventions that are beyond the control of individual physicians and patients. Therefore, our risk prediction assessment serves only as a guideline and should not be taken as an absolute definition of high risk.

We envisage that the chlamydia risk scoring tool developed in this study will be adapted for interactive 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 targeting those at higher risk for infection(s).

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.

Competing interests None.

Ethics approval Ethical Review Board of Maggiore Hospital - Bologna Local health district.

Contributors HW implemented the study, analysed the data and wrote the first draft. RG, BD and AM helped interpreting the data and finalising the manuscript. All authors saw and approved the final manuscript.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Jin F, Prestage GP, Zablotska I, *et al*. High rates of sexually transmitted infections in HIV positive homosexual men: data from two community based cohorts. *Sex Transm Infect* 2007;83:397–9.
2. Goller J, Gold J, Lim M, *et al*. Victorian Primary Care Network for sentinel surveillance on BBVs and STIs: an update. *Victorian Infectious Diseases Bulletin* 2007;10:37–9.
3. Katz B, Thom S, Blythe M, *et al*. Fertility in adolescent women previously treated for genitourinary chlamydial infection. *Adolesc Pediatr Gynecol* 1994;7:147–52.
4. Low N, Egger M, Sterne JA, *et al*. Incidence of severe reproductive tract complications associated with diagnosed genital chlamydial infection: the Uppsala Women's Cohort Study. *Sex Transm Infect* 2006;82:212–18.
5. Brocklehurst P, Rooney G. Interventions for treating genital chlamydia trachomatis infection in pregnancy. *Cochrane Database Syst Rev* 1998.
6. Allaire AD, Huddleston JF, Graves WL, *et al*. Initial and repeat screening for Chlamydia trachomatis during pregnancy. *Infect Dis Obstet Gynecol* 1998;6:116–22.
7. Bernstein KT, Marcus JL, Nieri G, *et al*. Rectal gonorrhea and chlamydia reinfection is associated with increased risk of HIV seroconversion. *J Acquir Immune Defic Syndr* 2010;53:537–43.
8. Vajdic CM, Middleton M, Bowden FJ, *et al*. The prevalence of genital Chlamydia trachomatis in Australia 1997–2004: a systematic review. *Sex Health* 2005;2:169–83.
9. Geisler W. Duration of untreated, uncomplicated Chlamydia trachomatis genital infection and factors associated with chlamydia resolution: a review of human studies. *J Infect Dis* 2010;201(Suppl 2): S104–13.
10. Rockhill B. The privatization of risk. *Am J Public Health* 2001;91:365–8.
11. Schulze MB, Hoffmann K, Boeing H, *et al*. An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care* 2007;30:510–15.
12. Vickers AJ, Basch E, Kattan MW. Against diagnosis. *Ann Intern Med* 2008;149:200–3.
13. Gonen M. *Analyzing Receiver Operating Characteristic Curves with SAS*. Cary, NC: SAS Institute, 2007.
14. Wand H, Spiegelman D, Law M, *et al*. Estimating population attributable risk for hepatitis C seroconversion in injecting drug users in Australia: implications for prevention policy and planning. *Addiction* 2009;104:2049–56.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Page 2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Page 3
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 3
Methods			
Study design	4	Present key elements of study design early in the paper	Page 3-4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 3-4
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Page 3-5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Page4-5
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 3-5
Bias	9	Describe any efforts to address potential sources of bias	Page 7-8
Study size	10	Explain how the study size was arrived at	Page 3
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Page 3-5
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Page 4-5
		(*b*) Describe any methods used to examine subgroups and interactions	Page 4
		(*c*) Explain how missing data were addressed	Page 4
		(*d*) If applicable, explain how loss to follow-up was addressed	N/A
		(*e*) Describe any sensitivity analyses	Page 6-7
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	Page 5
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Page 5-6
		(b) Indicate number of participants with missing data for each variable of interest	Page 4
		(c) Summarise follow-up time (eg, average and total amount)	N/A
Outcome data	15*	Report numbers of outcome events or summary measures over time	Page 5-6
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 5-6
		(*b*) Report category boundaries when continuous variables were categorized	Page 7
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Page 6
Discussion			
Key results	18	Summarise key results with reference to study objectives	Page 8
Limitations			
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Page 8
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 7
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 9

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

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 www.strobe-statement.org.