ORIGINAL PAPER

Risk Prediction Score for HIV Infection: Development and Internal Validation with Cross-Sectional Data from Men Who Have Sex with Men in China

Lu Yin[1,2] · Yuejuan Zhao[3] · Meridith Blevins Peratikos[4] · Liang Song[5] · Xiangjun Zhang[6] · Ruolei Xin[3] · Zheya Sun[7] · Yunan Xu[7,8] · Li Zhang[3] · Yifei Hu[7,9] · Chun Hao[10] · Yuhua Ruan[7] · Yiming Shao[7] · Sten H. Vermund[2,11] · Han-Zhu Qian[2,11]

Abstract

Receptive anal intercourse, multiple partners, condomless sex, sexually transmitted infections (STIs), and drug/alcohol addiction are familiar factors that correlate with increased human immunodeficiency virus (HIV) risk among men who have sex with men (MSM). To improve estimation to HIV acquisition, we created a composite score using questions from routine survey of 3588 MSM in Beijing, China. The HIV prevalence was 13.4%. A risk scoring tool using penalized maximum likelihood multivariable logistic regression modeling was developed, deploying backward step-down variable selection to obtain a reduced-form model. The full penalized model included 19 sexual predictors, while the reduced-form model had 12 predictors. Both models calibrated well; bootstrap-corrected c-indices were 0.70 (full model) and 0.71 (reduced-form model). Non-Beijing residence, short-term living in Beijing, illegal drug use, multiple male sexual partners, receptive anal sex, inconsistent condom use, alcohol consumption before sex, and syphilis infection were the strongest predictors of HIV infection. Discriminating higher-risk MSM for targeted HIV prevention programming using a validated risk score could improve the efficiency of resource deployment for educational and risk reduction programs. A valid risk score can also identify higher risk persons into prevention and vaccine clinical trials, which would improve trial cost-efficiency.

Keywords Men who have sex with men (MSM) · Risk score · HIV · HIV prevention · China

Introduction

Both cross-sectional and cohort studies have documented a rapid increase of HIV prevalence among men who have sex with men (MSM) in China (from 1.4% in 2001 to 6.3% in 2011) [1–3], which is especially notable in the nation's capital, Beijing [1, 4–9]. Multiple sexual partners, inconsistent condom use, and use of alcohol and aphrodisiac substances are associated with increasing HIV transmission risk

among Chinese MSM [1, 2, 5–8, 10–14]. MSM in Beijing have the option of visiting voluntary counseling and testing (VCT) clinics for HIV testing. Though MSM are advised to test every 6 months [15–17], frequent VCT among MSM in China is uncommon [18, 19] even in communities with good HIV risk awareness and social tolerance towards homosexuality [20–22]. Frequent testing could find those living with HIV as soon as possible, and subsequent ART treatment can reduce transmission to others. However, HIV-negative MSM are more likely to continue risk behaviors when they push their luck. One study cited an increase of high risk behaviors among HIV-negative MSM who tested for HIV regularly, so effective risk reduction counselling must be a part of VCT services [23]. Both governmental and non-governmental sources emphasize that the efforts and investments of interventions to reduce HIV transmission for MSM must be evaluated to see what works to curb the growing HIV epidemic [11, 24]. Recognizing MSM at highest risk of HIV/AIDS transmission can contribute to

---

Lu Yin, Yuejuan Zhao and Meridith Blevins Peratikos These authors contribute equally to this work.

✉ Sten H. Vermund
sten.vermund@yale.edu

Extended author information available on the last page of the article

more efficiently targeted interventions, either specific to or independent of Chinese sexual cultural context [25, 26].

Developing risk scores for one disease using clinical prediction rules can generate valuable public health tools [27–30]. Risk scores for HIV infection have used traditional variable selection of backward selection or stepwise procedures in maximum likelihood estimation [30–35], or manual variable selection and model comparisons based on the associations between HIV infection and each potential predictor [36, 37]. Identifying MSM at greatest risk for HIV infection can permit better to focus on prevention resources and interventions, such as enhanced counseling and/or antiretroviral pre-exposure prophylaxis (PrEP) [38, 39]. HIV prevention clinical trial designs can be made more efficient if higher risk persons are identified and recruited preferentially. We sought to develop a risk score for HIV infection based on Chinese MSM using Harrell's regression model strategies [40], which may reduce prediction error compared to traditional maximum likelihood estimation.

Methods

Study Population and Setting

Our NIH-funded combination HIV prevention project included two components: (1) promoting HIV testing in Beijing among MSM, including transgender women; and (2) a pilot randomized clinical trial of interventions for enhancing linkage to HIV care among newly diagnosed HIV-infected MSM. Data for this paper were drawn from the first component, a cross-sectional study. MSM participants were recruited mainly from the regular clients who visited four study clinics or through short message service, peer referral, web-advertisement and community outreach [41, 42]. Participant recruitment was conducted between March 2013 and March 2014 in four HIV VCT clinics in Beijing, including Beijing Municipal Center for Disease Control and Prevention (CDC), Xicheng District CDC, Chaoyang District CDC, and Jingcheng Skin Disease Hospital.

The inclusion criteria for study participation were: man or transgendered woman who self-reported having sex with a man in the last 12 months; age ≥ 18 years; living in Beijing; do not know their own HIV status or ever diagnosed as HIV negative in the last HIV testing (in order to find newly-diagnosed HIV positive MSM for second component of whole project); able and willing to provide written informed consent to questionnaire surveys and serological testing; and not having participated previously in HIV testing in this study. For identifying the latter, cell phone numbers of potential participants were used to screen for possible duplicate testing. This study protocol and informed consent form were reviewed and approved by the institutional review boards

of the National Center for AIDS/STD Control and Prevention (NCAIDS) of the Chinese Center for Disease Control and Prevention (FWA00002958) and Vanderbilt University Medical Center (FWA00005756).

Data Collection and Laboratory Tests

An interviewer-administered structured questionnaire was used to collect respondent potential risk predictors. All interviewers were trained of interview skills, protocol and questionnaires before study recruitment and all had prior work experience within HIV/AIDS-related programs. One private room was provided for face-to-face interview between one trained interviewer and each potential participant, which could prevent participants from telling a lie due to stigma. Negative results of both HIV and syphilis rapid tests using finger or venous blood at clinics were provided for participants following interview completion on the same day. Pre- and post-test counseling was provided to participants, including the significance and possible limitations of the test (e.g., HIV window periods) and safer sex practices. Further laboratory confirmatory tests of HIV and syphilis were performed within 2 weeks for positive screening test results. If venous blood was collected for rapid tests, the remaining of the blood sample was used for serological confirmatory tests. If a finger stick was used for rapid tests, and a positive result on either test was received, a venous blood sample was collected for confirmatory tests of both HIV and syphilis.

HIV rapid tests were performed using Alere Determine™ HIV1/2 (Colloidal Selenium Device; Alere Medical Co, Ltd., USA). Venous specimens were further screened for HIV antibody using ELISA (HIV ELISA testing kit 1, ZHUHAI LIVZON Diagnostics Inc., China). If positive by ELISA screening, the specimen was double tested using the same ELISA kit and another ELISA kit (HIV ELISA testing kit 2, Beijing WANTAI Biological Pharmacy Enterprise Co. Ltd., China). A specimen with a positive reaction in both or either ELISA testing kits was confirmed by Western blot test (HIV Blot 2.2 WB; MP Biomedicals Co, Ltd., China). One Step Anti-TP (Treponema pallidum/Syphilis) Test (Colloidal Gold Device, Standard Diagnostics, Inc., Korea) was used as a syphilis rapid test. For rapid test positive persons, syphilis serology was determined through Syphilis Toluidine red unheated serum test (Beijing WANTAI Biological Pharmacy Enterprise Co. Ltd., China) and confirmed by the Treponema palladium particle assay (Fujirebio Inc., Tokyo, Japan).

Candidate Variables

Potential predictors were considered for exposure to HIV risk (original questionnaire in Appendix S1) that could also be regularly collected among MSM in general HIV prevention and research settings. Nineteen candidate variables

were used based on plausibility of risk association (Appendix S2). The number of knots specified according the estimated "power" of each predictor. The ten candidate variables (items) were included in model as continuous variables using restricted cubic splines with four knots, which offer an adequate fit of the model and are a good compromise between flexibility and loss of precision caused by overfitting a small sample [40, 43]. If continuous data had low variability, the number of knots was reduced to 3, or continuous data were categorized as ordinal or dichotomous variables, where relevant. Default knot locations chose quantiles or tertiles of each predictor's distribution. Two items were collected using ordinal scales (Appendix S2).

Statistical Modeling and Sample Size Justification

The primary study outcome was HIV sero-status. The associations between potential predictors and HIV infection were evaluated using a logistic regression model. Sample size for prediction modeling must be large enough or have a conservative list of candidate predictors to achieve reproducibility in future populations. Specifically, 10 infections should be observed per parameter numbers estimated in the model. Our proposed model has 10 continuous variables, 6 dichotomous variables, and 3 ordinal variables that yield 40 parameters. With 455 HIV infections, we allowed 45 parameters to avoid overfitting.

Statistical Modeling of the Full Model and Reduced-form Model

The following analyses were conducted using R-software version 3.3.1 (*R* statistical software, Institute for Statistics and Mathematics, Vienna, Austria) [44, 45]. All pre-specified variables described above were included in a full logistic regression model of HIV infection using shrinkage of estimated regression coefficients, which might have larger predictive capability when used for future similar population. If selecting variables only through univariate screening and expert experiences, bias in estimating regression coefficients might be introduced [40, 46–48]. In order to correct for potential overfitting of regression coefficients, we applied penalized maximum likelihood estimation (PMLE), which maximized the log-likelihood of conventional logistic regression, provided smaller prediction errors, and preserved the discriminative ability of the model while shrinking each predictor for over-optimism [40, 43]. No predictor interactions were examined during modeling due to the complexity of developing and interpreting a model with interaction terms. To reduce the number of variables in the above-mentioned full model and to identify a subset of items that could be used in future surveys of assessing HIV infection risk among MSM, a reduced-form model was fitted using backward variable selection at an alpha criterion of 0.25 with bootstrap validation. Odds ratios (OR) 95% confidence intervals (CI), and nomograms were summarized for both the full and reduced-form models. Online evaluation tools were published using Harrell application by RStudio (http://shiny.rstudio.com/).

Model Calibration and Discrimination

Model accuracy regarding both calibration and discrimination was assessed [46, 49]. Calibration is the ability of our model to predict either future observations or the current survey data. Calibration was evaluated by graphing the predicted probabilities (x-axis) against observed outcomes (y-axis). Since observed outcomes are binary (1 = HIV-infected and 0 = not infected), we used locally weighted scatterplot smoothing (LOWESS) to estimate the observed probabilities of the outcome in relation to the predicted probabilities and compared this line with a 45° line which represents perfect calibration. Predictive discrimination was assessed using the c-index, which is a rank-order statistic to estimate the probability of concordance between the predicted and observed outcomes and to measure how well the model discriminates between different outcomes.

Internal Bootstrap Validation

Bootstrap validation was used to assess the predictive ability and reproducibility of the model [43, 46, 49, 50]. Internal bootstrap validation was performed as follows: (1) estimate the apparent predictive ability with the original sample dataset used to fit the model; (2) draw 300 bootstrap samples with replacement for fitting the model and measuring the predictive ability for each bootstrap sample; (3) test the accuracy measures by applying the bootstrap model to the original sample; (4) calculate the optimism in the predictive ability for each bootstrap model as the difference in performance between the bootstrap model and the original dataset; and (5) use the average of the optimism over all 300 bootstrap samples as a correction factor to estimate a validated performance measure. Thus, bootstrap-corrected estimates of the c-index were estimated for discrimination along with calibration curves for the full and reduced-form models [43].

Results

General Characteristics of Study Participants

During recruitment, 3751 MSM were screened for HIV infection. Fully 95.7% (n = 3,588) met inclusion criteria for our study and were used for HIV infection analyses. Participants (n = 163) were excluded due to the following

three reasons: (1) duplicate HIV testing using the same cell phone numbers in different study clinics or on different dates (n = 126); (2) diagnosed as HIV-infected prior to participation in this study since those who have knew their HIV positive status may change their behaviors (n = 30); (3) refusing venous or finger blood collection (n = 5); (4) incomplete questionnaire interview (n = 2).

The actual HIV prevalence was 13.4% among 3620 MSM after excluding 126 duplicate HIV testers and 5 refusing blood collection. Among 3588 eligible MSM who do not know their HIV status or HIV negative in the last HIV testing, and 12.7% tested positive for HIV. The mean age was 29.8 [standard deviation (SD), 7.8] years with the range of 18–75 years; 93.7% identified as Han ethnicity; 15.0% were currently married to a woman; 71.9% had obtained a bachelor' degree or higher; 80.6% were employed full time; 24.8% had Beijing permanent resident cards (called *Hukou*), without which MSM may have some restrictions, including buying apartment and cars, right to vote and to be elected, their children to school, and so on (protective impact of local residents on HIV risk was also observed in other Chinese big cities [1, 7, 51, 52]). Table 1 presents descriptive summaries of possible predictors and their associations with HIV infection estimated using univariate logistic regression models. Married currently with a woman, having Beijing *Hukou*, less years of living in Beijing, nitrite inhalants use (also called Rush), illegal drug use (including methamphetamine, MDMA, and ketamine), numbers of lifetime male sexual partners ≥ 10, unprotected receptive anal sex with male sexual partners, alcohol consumption often before sex, no sex with female sexual partners, consistent condom use with female sexual partners, and syphilis infection were associated with HIV acquisition (P < 0.05).

Full Model

HIV infection was modeled using penalized multivariable logistic regression with the 19 potential predictors (Table 2). Eleven variables were associated with HIV infection, including Beijing *Hukou* (yes vs. no; OR 0.55; 95% CI 0.38–0.82), years of living in Beijing (10 vs. 2 years; OR 0.66; 95% CI 0.51–0.86), illegal drug use in the past 3 months (yes vs. no; OR 4.34; 95% CI 2.40–7.87), alcohol use in the past 3 months (≤ 1 time/month vs. never; OR 1.36; 95% CI 1.04–1.77), alcohol consumption before sex in the past 3 months (often vs. never; OR 2.98; 95% CI 1.45–6.10), number of lifetime male sexual partners (20 vs. 4; OR 1.45; 95% CI 1.15–1.83), number of male sexual partners in the past 3 months (2 vs. 1; OR 0.84; 95% CI 0.72–0.98), frequency of receptive anal sex in the past 3 months (3 vs. 0; OR 1.75; 95% CI 1.30–2.35), proportion of condom use during receptive anal sex in the past 3 months (100% vs. 0%; OR 0.53; 95% CI 0.37–0.77), frequency of insertive

Table 1 Univariate associations of potential predictors with HIV infection risk

Predictors	HIV Prevalence % (N)	Odd ratio (95% confidence interval)	P value
HIV prevalence	12.7 (455/3588)		
Age at interview (year)			
Mean ± SD	29.8 ± 7.8	0.99 (0.98, 1.01)[a]	0.20
≤ 30	13.0 (301/2312)	Ref.	
> 30	12.1 (154/1276)	0.92 (0.75, 1.13)	0.41
Ethnicity			
Han	12.5 (421/3361)	Ref.	
Others	15.0 (34/227)	1.23 (0.84, 1.80)	0.28
Currently married with a woman			
No	13.2 (402/3049)	Ref.	
Yes	9.8 (53/539)	0.72 (0.53, 0.97)	0.03
Years of education			
≤ 12	12.4 (125/1009)	Ref.	
> 12	12.8 (330/2579)	1.04 (0.83, 1.29)	0.74
Full-time employed			
No	12.9 (90/695)	Ref.	
Yes	12.6 (365/2893)	0.97 (0.76, 1.24)	0.81
Beijing permanent residents (*Hukou*)			
No	14.4 (389/2699)	Ref.	
Yes	7.4 (66/889)	0.48 (0.36, 0.63)	< 0.01
Years of living in Beijing			
Median (IQR)	5 (2–10)	0.96 (0.93, 0.99)[a]	< 0.01
≤ 5	15.0 (278/1851)	Ref.	
> 5	10.2 (177/1737)	0.74 (0.58, 0.93)	0.01
Nitrite inhalants use in the past 3 months			
No	11.7 (307/2627)	Ref.	
Yes	15.4 (148/961)	1.38 (1.11, 1.70)	< 0.01
Illegal drug use in the past 3 months			
No	12.1 (428/3529)	Ref.	
Yes	45.8 (27/59)	6.11 (3.63, 10.30)	< 0.01
Alcohol use in the past 3 months			
Never	12.9 (203/1574)	Ref.	
≤ 1 time/month	10.9 (121/1108)	0.83 (0.65, 1.05)	0.12
2-4 times/month	14.6 (87/594)	1.16 (0.88, 1.52)	0.29
≥ 5 times/month	14.1 (44/312)	1.11 (0.78, 1.58)	0.56
Alcohol consumption before sex in the past 3 months			
Never	11.8 (337/2850)	Ref.	
Occasionally	14.8 (78/527)	1.30 (0.99, 1.69)	0.06
Sometimes	16.9 (25/148)	1.52 (0.97, 2.37)	0.07
Often	23.8 (15/63)	2.33 (1.29, 4.21)	0.01
Time since the first sex			
Median (IQR)	7 (4–12)	0.997 (0.98, 1.01)	0.68
< 8 years	12.3 (225/1833)	Ref.	
≥ 8 years	13.1 (230/1755)	0.93 (0.76, 1.23)	0.46
Time since the first anal sex			
Median (IQR)	6 (3–10)	0.996 (0.98, 1.01)	0.64
< 6 years	12.6 (208/1648)	Ref.	

Table 1 (continued)

Predictors	HIV Prevalence % (N)	Odd ratio (95% confidence interval)	P value
≥ 6 years	12.7 (247/1940)	1.01 (0.83, 1.23)	0.92
Number of lifetime male sexual partner			
Median (IQR)	9 (4–20)	1.00 (0.999, 1.001)	0.84
< 10	10.9 (198/1815)	Ref.	
≥ 10	14.5 (257/1773)	1.38 (1.14, 1.69)	< 0.01
Number of male sexual partners in the past 3 months			
Median (IQR1)	1 (1–2)	1.00 (0.98, 1.02)	0.73
≤ 1	11.8 (221/1879)	Ref.	
> 1	13.7 (234/1709)	1.19 (0.98, 1.45)	0.08
Unprotected receptive anal sex with male sexual partners in the past 3 months			
No	9.8 (268/2743)	Ref.	
Yes	22.1 (187/845)	2.63 (2.14, 3.22)	< 0.01
Unprotected insertive anal sex with male sexual partners in the past 3 months			
No	12.3 (334/2706)	Ref.	
Yes	13.7 (121/882)	1.13 (0.90, 1.41)	0.29
Sex with female partners in the past 3 months			
No	14.6 (150/1025)	Ref.	
Yes	11.9 (305/2563)	0.79 (0.64, 0.97)	0.03
Unprotected sex with female sexual partners in the past 3 months			
No	13.1 (438/3346)	Ref.	
Yes	4.0 (17/422)	0.50 (0.30, 0.83)	0.01
Syphilis infection			
No	11.4 (378/3319)	Ref.	
Yes	28.6 (77/269)	3.12 (2.35, 4.15)	< 0.01

SD standard deviation, *IQR* interquartile range

[a]Odds ratio per 1-unit increase

anal sex in the past 3 months (4 vs. 0; OR 0.70; 95% CI 0.52–0.93), and syphilis infection (yes vs. no; OR 2.87; 95% CI 2.11–3.91).

Reduced-form Model

Backward step-down variable selection of risk variables from the full model was used to obtain a reduced-form model for HIV infection (Table 2, Fig. 1, Appendices S3), including Beijing *Hukou* (yes vs. no; OR 0.55; 95% CI 0.38–0.81), years of living in Beijing (10 years vs. 2 years; OR 0.67; 95% CI 0.52–0.86), illegal drug use in the past 3 months (yes vs. no; OR 4.35; 95% CI 2.43–7.76), alcohol consumption before sex in the past 3 months (often vs. never; OR 2.83; 95% CI 1.48–5.39), time since the first sex (12 years vs. 4 years; OR 1.35; 95% CI 1.02–1.78), number of lifetime male sexual partners (20 vs. 4; OR 1.41; 95% CI 1.13–1.75), number of male sexual partners in the past 3 months (2 vs.

1; OR 0.85; 95% CI 0.73–0.99), frequency of receptive anal sex with male sexual partners in the past 3 months (3 vs. 0; OR 1.67; 95% CI 1.25–2.24), proportion of condom use during receptive anal sex with male sexual partners in the past 3 months (100% vs. 0%; OR 0.49; 95% CI 0.34–0.69), frequency of insertive anal sex with male partners in the past 3 months (4 vs. 0; OR 0.78; 95% CI 0.60–1.01), sex with female partners in the past 3 months (yes vs. no; OR 0.53; 95% CI 0.34–0.81), and syphilis infection (yes vs. no; OR 2.89; 95% CI 2.13–3.93). The reduced-form model yielded the bias-corrected c-indices of 0.71 (calibration curves in Appendices S4).

Nomogram

The nomogram for penalized reduced-form model of risk score (Fig. 1) may be more convenient to evaluate HIV predicted probability for MSM in practice. The *Points* grid-line at the top of the nomogram is assigned to point values for each sexual variable by aligning a vertical line from the gridline to the line adjacent to the sexual variable. The individual variable points are summed and a vertical line from the summed points on the *Total Points* grid-line to the *Predicted Probability* provides an estimated probability of HIV infection. For example, the corresponding probability of HIV infection is approximately 88.6% (95% CI, 74.2–95.4%) for a MSM who has no Beijing *Hukou*, lived in Beijing for 10 years before recruitment, illegal drug use, has 10 years of sexual experiences, has 20 male sexual partners in total and 5 male sexual partners in the past 3 months, 15 episodes of receptive anal sex with male sexual partners, 50% condom use during receptive anal sex with male partners, 5 episodes of insertive anal sex with male sexual partners, alcohol consumption often before sex, no sex with female partners in the past 3 months, and current syphilis infection. For sake of facilitating the calculation for a potential participant, our calculation is further illustrated using Shiny in Fig. 2 (available at https://meridithblevins.shinyapps.io/sexualrisk/).

Discussions

We have developed a discrete set of predictors for HIV risk among Chinese MSM who do not know their HIV status or HIV negative in the last HIV testing using Harrell's regression modeling strategies. The impact of each predictor might not be obvious on increasing HIV risk, but Harrell's regression includes several potential predictors together to obtain stronger prediction abilities [40]. The predictors selected for our reduced-form risk score model are independently associated with HIV infection and included variables commonly measured in HIV prevention studies among Chinese MSM. The reduced-form model seems to be more competitive than

Table 2 Full and reduced-form prediction rule models and penalized odds ratios for the associations of potential predictors with HIV risk

Potential items for sexual risk score	Type	df	PMLE model	
			Full model	Reduced-form model
Age at recruitment (change from 24 to 33 years)	Continuous	3	1.16 (0.79, 1.77)	n/a
Currently married with a woman	Dichotomous	1	0.80 (0.53, 1.21)	n/a
Beijing *Hukou*	Dichotomous	1	0.55 (0.38, 0.82)	0.55 (0.38, 0.81)
Years of living in Beijing (change from 2 to 10 years)	Continuous	2	0.66 (0.51, 0.86)	0.67 (0.52, 0.86)
Nitrite inhalants use in the past 3 months	Dichotomous	1	1.04 (0.82, 1.32)	n/a
Illegal drug use in the past 3 months	Dichotomous	1	4.34 (2.40, 7.87)	4.35 (2.43, 7.76)
Alcohol use in the past 3 months	Ordinal	3		
(≤ 1 time/month vs. never)			1.36 (1.04, 1.77)	n/a
(2–4 times/month vs. never)			1.33 (0.96, 1.84)	n/a
(≥ 5 times/month vs. never)			1.06 (0.68, 1.66)	n/a
Alcohol consumption before sex in the past 3 months	Ordinal	3		
(occasionally vs. never)			1.17 (0.85, 1.62)	1.13 (0.85, 1.50)
(sometimes vs. never)			1.29 (0.76, 2.19)	1.27 (0.78, 2.06)
(often vs. never)			2.98 (1.45, 6.10)	2.83 (1.48, 5.39)
Time since the first sex (change from 4 to 12 years)	Continuous	3	1.67 (0.95, 2.91)	1.35 (1.02, 1.78)
Time since the first anal sex (change from 3 to 10 years)	Continuous	3	0.75 (0.48, 1.18)	n/a
Number of lifetime male sexual partners (from 4 to 20)	Continuous	3	1.45 (1.15, 1.83)	1.41 (1.13, 1.75)
Number of male sexual partners in the past 3 months (from 1 to 2)	Continuous	3	0.84 (0.72, 0.98)	0.85 (0.73, 0.99)
Frequency of receptive anal sex in the past 3 months (from 0 to 3)	Continuous	2	1.75 (1.30, 2.35)	1.67 (1.25, 2.24)
Proportion of condom use during receptive anal sex with male partners in the past 3 months (from 0 to 100%)	Continuous	3	0.53 (0.37, 0.77)	0.49 (0.34, 0.69)
Frequency of insertive anal sex in the past 3 months (from 0 to 4)	Continuous	2	0.70 (0.52, 0.93)	0.78 (0.60, 1.01)
Proportion of condom use during insertive anal sex with male partners in the past 3 months (from 0 to 100%)	Continuous	3	0.70 (0.48, 1.02)	n/a
Sex with female partners in the past 3 months	Dichotomous	1	0.56 (0.27, 1.14)	0.53 (0.34, 0.81)
Condom use during sex with female partners in the past 3 months	Ordinal	2		
(sometimes versus consistent)			0.77 (0.22, 2.67)	n/a
(never vs. consistent)			0.69 (0.21, 2.30)	n/a
Syphilis infection	Dichotomous	1	2.87 (2.11, 3.91)	2.89 (2.13, 3.93)
Original c-index	Total df	40	0.73	0.72
Bias-corrected c-index			0.70	0.71

df degree of freedom, *PMLE* penalized maximum likelihood estimation, *UOR* unadjusted odds ratio, *AOR* adjusted odds ratio, *CI* confidence interval, *n/a* not available

the full model for HIV infection among Chinese MSM, and corresponding nomogram using Shiny is more convenient for health workers or education counsellors during routine prevention and intervention. To our knowledge, this is the first risk score derived from Chinese MSM, obtaining bias-corrected c-index of 0.71 in reduced-form model. Some risk scores have been developed and further simplified to predict the HIV infection probability showing c-index from 0.66 to 0.74 among American MSM [32, 33, 35, 53].

As we know, prediction errors and over-optimism derived from traditional regression models are inevitable. In addition, categorizing continuous variables might produce arbitrary cutoff values for some predictors when they were used in other studies. Harrell's strategies utilizing PMLE, restricted cubic splines, and LOWESS have been confirmed to improve prediction capability of the model [40, 50], and used successfully to predict pediatric acute asthma for hospitalization [43]. In order to facilitate calculating the predicted probability of HIV infection for each MSM, our online risk tool is provided for clinic counsellors and physicians. A high predicted probability of HIV infection could influence the efficient allocation of prevention resources and careful selection of appropriate interventions after comprehensive consideration and assessment for the balance of benefits versus potential toxicity, especially enhanced counselling and PrEP [38, 39].

Although our risk score was derived from a cross-sectional study and validated internally using 3588 MSM in Beijing, China, external validation to MSM in other regions
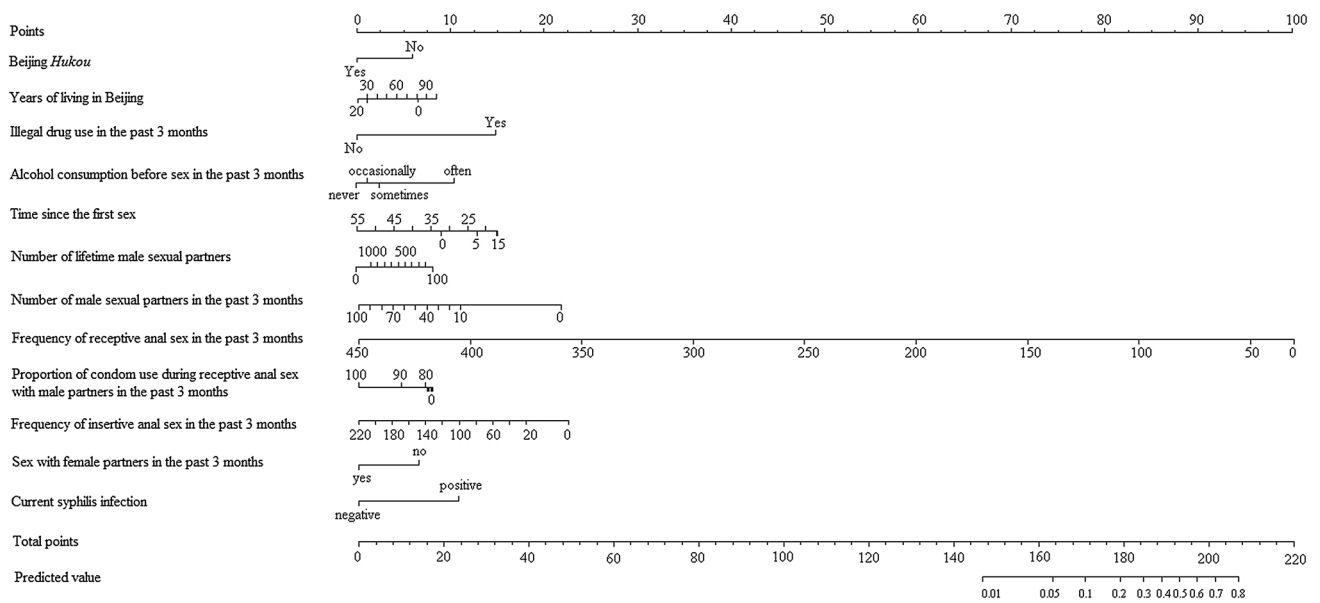
Fig. 1 Reduced-form sexual risk score PMLE model nomogram for HIV infection

Fig. 2 Screenshot of calculating predicted probability for each MSM using Shiny app

should be considered before widespread deployment. We did not develop this score to differentiate MSM in the process of pre- and post-counselling because a low score did not indicate zero HIV risk in our analyses. All MSM should be counselled about risk reduction strategies, including consistent condom use, the treatment of sexually transmitted infections (STIs), and exclusive relationships with sexual partners, regardless of female or male sexual partners. We think that the greatest utility of the scores will be in repeated assessments to track changes in HIV risk among MSM at set time intervals (e.g., every 3 or 6 months), helping to provide feedback for health workers or counsellors alike.

It may also help identify sub-populations with a need for more intense intervention, as well as helping identify highest risk persons for HIV prevention clinical trial recruitment. Additionally, definitive HIV test confirmation delays the final determination of sero-status. For example, in China it commonly takes 1–2 weeks to receive Western Blot results. Some HIV-infected patients are sometimes not linked to care after diagnosis due to this notification delay or due to a lack of identifiable information. Using this score evaluation might help health workers identify high risk MSM, whose identity and contact information could be collected at clinic even before infection status is confirmed.

The HIV prevalence in our study (13.4%) is no doubt concerning as it is much higher than previous studies among MSM in Beijing [1, 3–9], which was consistent with Beijing Sentinel Surveillance on the same year (unpublished data). The increasing HIV epidemic might be substantially worse due to an increase of testing coverage provided by various large testing programs targeting MSM using significant monetary incentives [1, 2]. In addition, WHO issued a new guideline in July 2013 recommending earlier treatment and raised the threshold for starting ART from a CD4 of 350 cells/μL to 500 cells/μL [54], and Chinese government guidelines also loosened CD4 criteria for ART initiation at the end of 2013. Hence, the number of people living with HIV are increasing substantially which has stimulated great debate over disease burden and treatment costs in China.

It is worth mentioning that those with more alcohol use were not likely to have higher acquisition risk in our MSM participants like previous studies [12, 14, 55]. However, alcohol use before sex showed a positive dose–response association with higher acquisition risk in our study, which also was included in reduced-form model. Additionally, heavy alcohol use is not very common among Chinese MSM [12, 14, 41, 55]. Though the positive association between heavy alcohol use and HIV infection was reported in our study [41], heavy alcohol use defined by alcohol use disorder identification test-consumption (AUDIT-C) was not used in most MSM studies so far. Therefore, AUDIT-C was not considered in developing our prediction model.

Our analysis has specific limitations. First, all questionnaire data were based on self-reporting; study participants may have provided responses based more on social desirability, especially involving sensitive questions, such as sexual risk behaviors. Second, bias or spurious associations might be introduced because our participants in this study are volunteers recruited using non-randomized sampling, who instead felt they had recently engaged in risky behaviors or could differ on any number of other factors. Hence, the study sample may not represent the entire MSM community in Beijing or in China. Third, our risk score was validated only using our study, and further validation should be completed using data from other MSM studies in other locations prior

to widespread implementation [27, 28, 56, 57]. Fourth, we simply measured prevalence at a single point in time among MSM of varying ages who happened to seek testing at our VCT clinics. Many variables included in the model were time-varying, such as age at the time of recruitment and number of lifetime male sexual partners, meaning that they would be expected to change throughout the course of an individual's life thereby changing their predicted probability of HIV infection. Additionally, correlating a time-varying variable with a cross-sectional HIV test isn't valid because nobody knows how many lifetime male sexual partnerships an individual had at the time he acquired HIV. Fifth, the recall period of recent risky sexual behaviors was set in the past 3 months in our study. Moderator numbers of Chinese MSM may change short-term sexual behaviors due to personal reasons, including feeling sick, career promotion, job change, and so on. Longer recall period might be preferred to find risk behaviors, since our study design is unable to collect risky behaviors when they infected. Our prediction model might also be altered a little if this recall period was increased to 6 months or decreased to 1 month in other studies. Finally, our cross-sectional study design has potential methodological limitations to detect the causality between exposure and outcome.

Conclusions

It remains challenging to provide early diagnosis, linkage to care, and early treatment to HIV-infected individuals. A simple, quantitative risk score could provide a robust and useful approach to discriminate MSM at highest risk for HIV acquisition and to evaluate changes of HIV risk over time in clinical, public health, or research settings. Use of a validated risk score may also help improve research efficiency in prevention studies such as PrEP by discriminating higher from lower risk persons.

Compliance with Ethical Standards

Conflict of interest All co-authors have no conflict of interest.

Ethical Approval This article does not contain any studies with animals performed by any of the authors. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

1. Wu Z, Xu J, Liu E, Mao Y, Xiao Y, Sun X, et al. HIV and syphilis prevalence among men who have sex with men: a cross-sectional survey of 61 cities in China. Clin Infect Dis. 2013;57(2):298–309.

2. Chow EP, Lau JT, Zhuang X, Zhang X, Wang Y, Zhang L. HIV prevalence trends, risky behaviours, and governmental and community responses to the epidemic among men who have sex with men in China. Biomed Res Int. 2014;2014:607261.

3. Chow EP, Wilson DP, Zhang J, Jing J, Zhang L. Human immunodeficiency virus prevalence is increasing among men who have sex with men in China: findings from a review and meta-analysis. Sex Transm Dis. 2011;38(9):845–57.

4. Ma X, Zhang Q, He X, Sun W, Yue H, Chen S, et al. Trends in prevalence of HIV, syphilis, hepatitis C, hepatitis B, and sexual risk behavior among men who have sex with men. Results of 3 consecutive respondent-driven sampling surveys in Beijing, 2004 through 2006. J Acquir Immune Defic Syndr. 2007;45(5):581–7.

5. Ruan Y, Jia Y, Zhang X, Liang H, Li Q, Yang Y, et al. Incidence of HIV-1, syphilis, hepatitis B, and hepatitis C virus infections and predictors associated with retention in a 12-month follow-up study among men who have sex with men in Beijing, China. J Acquir Immune Defic Syndr. 2009;52(5):604–10.

6. Li S, Zhou Z, Jiang S, Liu Y, Li D, Zhang Z, et al. Incidence and risk factors of HIV and syphilis seroconversion among men who have sex with men in Beijing. Zhonghua Yu Fang Yi Xue Za Zhi. 2011;45(2):118–22 **[Chinese]**.

7. Fan S, Lu H, Ma X, Sun Y, He X, Li C, et al. Behavioral and serologic survey of men who have sex with men in Beijing, China: implication for HIV intervention. AIDS Patient Care STDS. 2012;26(3):148–55.

8. Li D, Li S, Liu Y, Gao Y, Yu M, Yang X, et al. HIV incidence among men who have sex with men in Beijing: a prospective cohort study. BMJ Open. 2012;2(6):e001829. https://doi.org/10.1136/bmjopen-2012-001829.

9. Li D, Gao Y, Yu M, Yang X, Li S, Xu J, et al. Study on the incidence of HIV and associated risk factors through a prospective cohort among men who have sex with men in Beijing, China. Zhonghua Liu Xing Bing Xue Za Zhi. 2012;33(7):663–6 **[Chinese]**.

10. Hao C, Yan H, Yang H, Huan X, Guan W, Xu X, et al. The incidence of syphilis, HIV and HCV and associated factors in a cohort of men who have sex with men in Nanjing, China. Sex Transm Infect. 2011;87(3):199–201.

11. Lau JT, Lin C, Hao C, Wu X, Gu J. Public health challenges of the emerging HIV epidemic among men who have sex with men in China. Public Health. 2011;125(5):260–5.

12. Lu H, Han Y, He X, Sun Y, Li G, Li X, et al. Alcohol use and HIV risk taking among Chinese MSM in Beijing. Drug Alcohol Depend. 2013;133(2):317–23.

13. Li D, Yang X, Zhang Z, Qi X, Ruan Y, Jia Y, et al. Nitrite inhalants use and HIV infection among men who have sex with men in China. Biomed Res Int. 2014;2014:365261.

14. Liu Y, Qian HZ, Ruan Y, Yin L, Ma J, Dahiya K, et al. Alcohol use among Chinese men who have sex with men: an epidemiological survey and meta-analysis. Biomed Res Int. 2014;2014:414381.

15. WHO. UNAIDS. Guidance on provider-initiated HIV testing and counselling in health facilities. Geneva: WHO Press; 2007.

16. WHO. UNODC. Guidance on testing and counselling for HIV in settings attended by people who inject drugs: improving access to treatment, care and prevention. Geneva: WHO Press; 2009.

17. World Health Organization, UNAIDS. Guidance on provider-initiated HIV testing and counselling in health facilities. Switzerland. (http://www.who.int/hiv/pub/guidelines/9789241595568_en.pdf) 2007.

18. Wei C, Ruan S, Zhao J, Yang H, Zhu Y, Raymond HF. Which Chinese men who have sex with men miss out on HIV testing? Sex Transm Infect. 2011;87(3):225–8.

19. Zou H, Hu N, Xin Q, Beck J. HIV testing among men who have sex with men in China: a systematic review and meta-analysis. AIDS Behav. 2012;16(7):1717–28.

20. Wu Z, Sun X, Sullivan SG, Detels R. Public health. HIV testing in China. Science. 2006;312(5779):1475–6.

21. Fan W, Yin L, Qian HZ, Li D, Shao Y, Vermund SH, et al. HIV risk perception among HIV negative or status-unknown men who have sex with men in China. Biomed Res Int. 2014;2014:232451.

22. Li X, Lu H, Raymond HF, Sun Y, Jia Y, He X, et al. Untested and undiagnosed: barriers to HIV testing among men who have sex with men, Beijing, China. Sex Transm Infect. 2012;88(3):187–93.

23. Lau JT, Li D, Wang Z, Lai CH. Repeated HIV voluntary counseling and testing increased risk behaviors among men who have sex with men in China: a prospective cohort study. AIDS Behav. 2015;19(11):1966–77.

24. Lu H, Liu Y, Dahiya K, Qian HZ, Fan W, Zhang L, et al. Effectiveness of HIV risk reduction interventions among men who have sex with men in China: a systematic review and meta-analysis. PLoS ONE. 2013;8(8):e72747.

25. Liao W, Lau JT, Tsui HY, Gu J, Wang Z. Relationship between sexual compulsivity and sexual risk behaviors among Chinese sexually active males. Arch Sex Behav. 2015;44(3):791–8.

26. Schneider JA, Zhou AN, Laumann EO. A new HIV prevention network approach: sociometric peer change agent selection. Soc Sci Med. 2015;125:192–202.

27. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. JAMA. 2000;284(1):79–84.

28. Shapiro SE. Guidelines for developing and testing clinical decision rules. West J Nurs Res. 2006;28(2):244–53.

29. Chaudhury R, Jones HE, Wechsberg W, O'Grady KE, Tuten M, Chisolm MS. Addiction severity index composite scores as predictors for sexual-risk behaviors and drug-use behaviors in drug-using pregnant patients. Am J Drug Alcohol Abuse. 2010;36(1):25–30.

30. Kahle EM, Hughes JP, Lingappa JR, John-Stewart G, Celum C, Nakku-Joloba E, et al. An empiric risk scoring tool for identifying high-risk heterosexual HIV-1-serodiscordant couples for targeted HIV-1 prevention. J Acquir Immune Defic Syndr. 2013;62(3):339–47.

31. Powers KA, Miller WC, Pilcher CD, Mapanje C, Martinson FE, Fiscus SA, et al. Improved detection of acute HIV-1 infection in sub-Saharan Africa: development of a risk score algorithm. Aids. 2007;21(16):2237–42.

32. Menza TW, Hughes JP, Celum CL, Golden MR. Prediction of HIV acquisition among men who have sex with men. Sex Transm Dis. 2009;36(9):547–55.

33. Smith DK, Pals SL, Herbst JH, Shinde S, Carey JW. Development of a clinical screening index predictive of incident HIV infection among men who have sex with men in the United States. J Acquir Immune Defic Syndr. 2012;60(4):421–7.

34. Wahome E, Fegan G, Okuku HS, Mugo P, Price MA, Mwashigadi G, et al. Evaluation of an empiric risk screening score to identify acute and early HIV-1 infection among MSM in Coastal Kenya. AIDS. 2013;27(13):2163–6.

35. Hoenigl M, Weibel N, Mehta SR, Anderson CM, Jenks J, Green N, et al. Development and validation of the San Diego Early Test Score to predict acute and early HIV infection risk in men who have sex with men. Clin Infect Dis. 2015;61(3):468–75.

36. Haukoos JS, Lyons MS, Lindsell CJ, Hopkins E, Bender B, Rothman RE, et al. Derivation and validation of the Denver Human Immunodeficiency Virus (HIV) risk score for targeted HIV screening. Am J Epidemiol. 2012;175(8):838–46.

37. Hsieh YH, Haukoos JS, Rothman RE. Validation of an abbreviated version of the Denver HIV risk score for prediction of HIV infection in an urban ED. Am J Emerg Med. 2014;32(7):775–9.

38. Centers for Disease Control and Prevention. (CDC). Interim guidance: preexposure prophylaxis for the prevention of HIV infection in men who have sex with men. MMWR Morb Mortal Wkly Rep. 2011;60(3):65–8.

39. Buchbinder SP, Glidden DV, Liu AY, McMahan V, Guanira JV, Mayer KH, et al. HIV pre-exposure prophylaxis in men who have sex with men and transgender women: a secondary analysis of a phase 3 randomised controlled efficacy trial. Lancet Infect Dis. 2014;14(6):468–75.

40. Harrell FE Jr. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Berlin: Springer; 2001.

41. Liu Y, Ruan Y, Strauss SM, Yin L, Liu H, Amico KR, et al. Alcohol misuse, risky sexual behaviors, and HIV or syphilis infections among Chinese men who have sex with men. Drug Alcohol Depend. 2016;168:239–46.

42. Liu Y, Vermund SH, Ruan Y, Liu H, Zhang C, Yin L, et al. HIV testing and sexual risks among migrant men who have sex with men: findings from a large cross-sectional study in Beijing, China. AIDS Care. 2017. https://doi.org/10.1080/09540121.2017.1381331.

43. Arnold DH, Gebretsadik T, Moons KG, Harrell FE, Hartert TV. Development and internal validation of a pediatric acute asthma prediction rule for hospitalization. J Allergy Clin Immunol Pract. 2015;3(2):228–35.

44. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Core Team. http://www.r-project.org/. Accessed 9 Mar 2015.

45. Harrell FE, Jr. The rms Package for R: regression modeling strategies. Nashville, USA. http://biostat.mc.vanderbilt.edu/wiki/Main/Rrms2015. Accessed 15 Feb 2015.

46. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361–87.

47. Steyerberg EW, Eijkemans MJ, Habbema JD. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol. 1999;52(10):935–42.

48. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Med Decis Making. 2001;21(1):45–56.

49. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130(6):515–24.

50. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer; 2009.

51. Wu X, Hong F, Lan L, Zhang C, Feng T, Yang Y. Poor awareness of syphilis prevention and treatment knowledge among six different populations in south China. BMC Public Health. 2016;16:287.

52. Ruan S, Yang H, Zhu Y, Wang M, Ma Y, Zhao J, et al. Rising HIV prevalence among married and unmarried among men who have sex with men: Jinan, China. AIDS Behav. 2009;13(4):671–6.

53. Scott H, Vittinghoff E, Irving R, Liu A, Fields SD, Magnus M, et al. Sex pro: a personalized HIV risk assessment tool for men who have sex with men. In: Conference on Retroviruses and Opportunistic Infections (CROI); Seattle, Washington. 2015.

54. WHO. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. Kuala Lumpur, Malaysia. June 2013.

55. Liao M, Kang D, Tao X, Bouey JH, Aliyu MH, Qian Y, et al. Alcohol use, stigmatizing/discriminatory attitudes, and HIV high-risk sexual behaviors among men who have sex with men in China. Biomed Res Int. 2014;2014:143738.

56. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. N Engl J Med. 1985;313(13):793–9.

57. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. JAMA. 1997;277(6):488–94.

Affiliations

Lu Yin[1,2] · Yuejuan Zhao[3] · Meridith Blevins Peratikos[4] · Liang Song[5] · Xiangjun Zhang[6] · Ruolei Xin[3] · Zheya Sun[7] · Yunan Xu[7,8] · Li Zhang[3] · Yifei Hu[7,9] · Chun Hao[10] · Yuhua Ruan[7] · Yiming Shao[7] · Sten H. Vermund[2,11] · Han-Zhu Qian[2,11]

[1] Medical Research & Biometrics Center, National Center for Cardiovascular Diseases, Beijing, China

[2] Vanderbilt Institute for Global Health, Vanderbilt University Medical Center, Nashville, TN, USA

[3] Beijing Municipal Center for Disease Control and Prevention & Centers for Preventive Medical Research, Beijing, China

[4] Departments of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA

[5] Chaoyang District Center for Disease Control and Prevention, Beijing, China

[6] Xicheng District Center for Disease Control and Prevention, Beijing, China

[7] State Key Laboratory of Infectious Disease Prevention and Control, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Chinese Center for Disease Control and Prevention, Beijing, China

[8] Department of Epidemiology, College of Public Health and Health Professions, University of Florida, Gainesville, FL, USA

[9] Department of Child, Adolescent Health and Maternal Health, School of Public Health, Capital Medical University, Beijing, China

[10] School of Public Health, Sun Yat-sen University, Guangzhou, China

[11] Yale School of Public Health, Yale University, New Haven, CT, USA