



Practice of Epidemiology

Derivation and Validation of the Denver Human Immunodeficiency Virus (HIV) Risk Score for Targeted HIV Screening

Jason S. Haukoos*, Michael S. Lyons, Christopher J. Lindsell, Emily Hopkins, Brooke Bender, Richard E. Rothman, Yu-Hsiang Hsieh, Lynsay A. MacLaren, Mark W. Thrun, Comilla Sasson, and Richard L. Byyny

* Correspondence to Dr. Jason S. Haukoos, Department of Emergency Medicine, Denver Health Medical Center, 777 Bannock Street, Mail Code 0108, Denver, CO 80204 (e-mail: jason.haukoos@dhha.org).

Initially submitted May 5, 2011; accepted for publication September 29, 2011.

Targeted screening remains an important approach to human immunodeficiency virus (HIV) testing. The authors aimed to derive and validate an instrument to accurately identify patients at risk for HIV infection, using patient data from a metropolitan sexually transmitted disease clinic in Denver, Colorado (1996–2008). With multivariable logistic regression, they developed a risk score from 48 candidate variables using newly identified HIV infection as the outcome. Validation was performed using an independent population from an urban emergency department in Cincinnati, Ohio. The derivation sample included 92,635 patients; 504 (0.54%) were diagnosed with HIV infection. The validation sample included 22,983 patients; 168 (0.73%) were diagnosed with HIV infection. The final score included age, gender, race/ethnicity, sex with a male, vaginal intercourse, receptive anal intercourse, injection drug use, and past HIV testing, and values ranged from -14 to $+81$. For persons with scores of <20 , $20-29$, $30-39$, $40-49$, and ≥ 50 , HIV prevalences were 0.31% (95% confidence interval (CI): 0.20, 0.45) ($n = 27/8,782$), 0.41% (95% CI: 0.29, 0.57) ($n = 36/8,677$), 0.99% (95% CI: 0.63, 1.47) ($n = 24/2,431$), 1.59% (95% CI: 1.02, 2.36) ($n = 24/1,505$), and 3.59% (95% CI: 2.73, 4.63) ($n = 57/1,588$), respectively. The risk score accurately categorizes patients into groups with increasing probabilities of HIV infection.

decision support techniques; HIV; HIV infection; prediction; prevalence; validation

Abbreviations: CDC, Centers for Disease Control and Prevention; HIV, human immunodeficiency virus; ROC, receiver operating characteristic; STD, sexually transmitted disease.

In 2006, the Centers for Disease Control and Prevention (CDC) published revised recommendations for performing human immunodeficiency virus (HIV) testing in health-care settings in the United States, specifically suggesting the use of routine (nontargeted) opt-out HIV screening in all health-care settings where the undiagnosed prevalence is greater than or equal to 0.1% (1). The rationale for this approach included identifying more patients with HIV infection and identifying them earlier in the course of their disease, thus helping to mitigate individual morbidity and transmission of the virus (2). Unfortunately, nontargeted HIV screening has not been widely adopted in clinical practice (3–5), partly because it is operationally difficult (6) and costly (7, 8) and requires screening of a large number of persons

in order to identify a modest number of newly diagnosed patients, in both high-prevalence and low-prevalence settings (7, 9).

In 2007, the World Health Organization and the US Preventive Services Task Force published recommendations for performing HIV testing in health-care settings, but contrary to the CDC recommendations, they endorsed targeted screening as the primary means of HIV testing (10, 11). These recommendations were put forth, in part, because of the paucity of data to support nontargeted screening in settings where the epidemic remains low-level or concentrated, including North America, Europe, and segments of other continents where a wide variation in the epidemiologic profile of HIV still exists (10, 12).

Although the concept of targeted HIV screening has existed for approximately 25 years and risk characteristics have been widely studied (13–15), specific targeted screening approaches remain largely undefined (16–19). Specifically, it remains unclear which criteria should be used to target patients, what the relative strengths of their associations are with HIV infection, and how they may be combined and incorporated into clinical practice to target an unselected population. A critical first step, however, is rigorous assessment of patient characteristics and their associations with HIV diagnosis. It would be useful to develop a tool that could be used to objectively estimate a patient's risk of having undiagnosed HIV infection. As such, our goal in this study was to systematically evaluate a large number of characteristics and to derive and validate a clinically meaningful prediction tool with which to accurately categorize patients into risk groups for undiagnosed HIV infection.

MATERIALS AND METHODS

Derivation of the Denver HIV risk score

Study design, setting, and population. We conducted an analysis of a large, prospectively collected data set from the Denver Metro Health Clinic, a sexually transmitted disease (STD) clinic administered by Denver Public Health in Denver, Colorado. The clinic is the largest of its kind in the Rocky Mountain region and serves over 10,000 patients annually, with an undiagnosed HIV prevalence of approximately 0.5%.

Consecutive patients aged 13 years or older who came to the STD clinic between January 1, 1996, and December 31, 2008, were included in the derivation sample. All patients visiting the STD clinic for outpatient HIV testing underwent structured health and behavioral screening as part of traditional HIV prevention counseling. Information on patient characteristics was collected by trained clinic staff using a computerized data collection system, allowing for real-time data entry.

As part of routine assessment, the following data were collected from each patient prior to testing: 1) demographic characteristics (age, gender, and race/ethnicity); 2) symptoms (rash, pruritus, genital discharge, and dysuria); 3) history of sexually transmitted infections (gonorrhea, chlamydia, herpes simplex, syphilis, and genital warts); 4) sexual history (lifetime number of partners and numbers of partners in the previous month and the previous 4 months); 5) specific sexual practices (vaginal intercourse, insertive or receptive oral intercourse, insertive or receptive anal intercourse, and condom use); 6) gender(s) of sexual contacts; 7) previous HIV testing history; and 8) other risk factors associated with the transmission of HIV infection (injection drug use, prostitution, or sexual contact with a prostitute or someone who injects drugs or is infected with HIV). All of these variables were available for analysis and served as candidate predictor variables during the development of the risk model. All patients underwent either conventional or rapid HIV testing, and all who tested preliminarily positive completed Western blot confirmatory testing. Confirmed HIV infection served as the outcome, or dependent variable.

Model development. Our modeling approach proceeded in a systematic fashion and included several sequential steps that are detailed in the Web Appendix (<http://aje.oxfordjournals.org/>) (20–22). We used multivariable logistic regression to develop the prediction model. Bivariate statistical testing or automated variable selection techniques were specifically not used to drive selection of variables in the model (23, 24). Instead, variables were included in the model using several sequential manual steps based on our knowledge of the epidemiology of HIV and known or hypothesized associations between patient characteristics and HIV infection. Variables were selected for inclusion based on their statistical associations, their influence on the regression coefficients of other included variables, and Akaike's Information Criterion as a global test of model goodness of fit.

The first step consisted of including only patient demographic characteristics in the model. Because age was a continuous variable and its relation with HIV infection was expected to be nonlinear, we used fractional polynomials to model its relation with the outcome. Fractional polynomials evaluate sets of 1- and 2-order transformations using a limited set of potential transformations of a variable relative to the outcome. From these sets, a "best"-fit transformation is identified. This approach is considered superior to modeling of continuous data as linear or arbitrarily identifying cut-points within the variable, because it accounts for potential nonlinear relations between the predictor and the outcome (25, 26). Using our results, we graphed the relation between age and HIV diagnosis to identify potential inflection points by which the probability of HIV diagnosis changed. As a result of this analysis, we categorized age into the following 4 mutually exclusive groups: <22 or >60 years, 22–25 or 55–60 years, 26–32 or 47–54 years, and 33–46 years.

Subsequent steps included the addition of variables related to sexual history, sexual orientation, sexual practices, other risk behaviors, history of sexually transmitted infections, and clinical symptoms. These variables were entered manually into the model in sequential groups. Combinations of variables were also evaluated (e.g., insertive and receptive oral intercourse combined as insertive or receptive oral intercourse). A full model was developed to include all statistically significant variables defined by a *P* value of 0.05 or less. Interaction terms were not evaluated for inclusion in the model given the large number of candidate variables and the complexity of developing and interpreting a model with interaction terms.

The internal validity of the model was assessed using 10-fold cross-validation. Calibration was assessed by graphically comparing predicted HIV prevalence with observed HIV prevalence, drawing a linear regression line through the points, and calculating its slope and R^2 . Discrimination was assessed by constructing a receiver operating characteristic (ROC) curve and calculating the area under the curve (27).

We then pruned the full model to develop the simplest model without losing its predictive ability. We retained the demographic variables and individually removed other risk score variables, beginning with those that had the weakest associations with HIV infection. Cross-validation was performed after removal of each variable, and new calibration

Table 1. Patient Characteristics From the Derivation Cohort ($n = 92,635$) Used as Potential Predictor Variables During Multivariable Modeling, Denver, Colorado, 1996–2008

	Diagnosis of HIV						Missing Data	
	Yes ($n = 504$)			No ($n = 92,131$)			No.	%
	No.	%	Median (IQR)	No.	%	Median (IQR)		
Demographic factors								
Age, years	32		26–38	27		22–35	0	0
Male gender	430	85		56,520	61		0	0
Race/ethnicity								
Black	119	24		22,620	25		0	0
Hispanic	129	26		26,652	29		0	0
White	241	48		38,928	42		0	0
Other ^a	15	3		3,931	4		0	0
Sexual history								
Lifetime no. of partners	20		7–50	10		6–25	6,399	7
No. of partners in previous 4 months	2		1–4	1		1–2	124	0.1
No. of partners in previous month	0		0–1	0		0–1	133	0.1
Sexual orientation								
Male heterosexual	96	19		48,647	53		68	0.07
Female heterosexual	61	12		31,429	34		58	0.06
Male who has sex with males	137	27		2,540	3		66	0.7
Female who has sex with females	1	0		397	0		66	0.7
Male bisexual	195	39		5,098	6		168	0.2
Female bisexual	11	2		3,717	4		159	0.2
Sexual practices								
Insertive oral intercourse	342	68		49,592	54		0	0
Receptive oral intercourse	360	71		55,893	61		8	0.009
Insertive anal intercourse	218	43		8,637	9		0	0
Receptive anal intercourse	234	46		7,135	8		8	0.009
Vaginal intercourse	178	35		80,334	87		21	0.02
Condom use	331	66		52,295	57		149	0.2

Table continues

and ROC curves were generated and compared with those of the preceding model. This pruning approach continued until the simplest model was created while ensuring maximal calibration and discrimination. A risk score was then created by multiplying the final model's regression coefficients by 10 and rounding them to the nearest integer (28). The range of the risk scores was then examined to identify cutpoints for grouping patients into statistically unique categories based on HIV prevalence. Proportions are reported with 95% confidence intervals.

Sample size estimation. We estimated the necessary sample size for the derivation of the risk model to minimize the possibility of overfitting the multivariable logistic regression model. As a general rule, multivariable logistic regression analyses require an event (i.e., HIV infection)-to-predictor ratio of 10:1 (29). We evaluated approximately 50 candidate predictor variables, so we estimated that we would require at least 500 newly diagnosed HIV-infected patients in the derivation sample. Given an estimated prevalence of HIV infection from the STD clinic of 0.5%, we estimated that we

would require approximately 100,000 patients in total for the derivation part of this study.

Validation of the Denver HIV risk score

Study design, setting, and population. The derived risk score was then externally tested using observations from January 1, 1998, through June 30, 2010, from the emergency department at the University of Cincinnati Medical Center in Cincinnati, Ohio. The University of Cincinnati Medical Center is a 665-bed tertiary-care facility and level 1 trauma center. There are approximately 90,000 adult visits to the emergency department annually. A targeted emergency department HIV screening program, in continuous operation since 1998, is an adjunct clinical program of the emergency department and is staffed by trained counselors who provide testing (using conventional HIV enzyme immunoassay with confirmatory Western blot) and comprehensive prevention counseling (30). Patients are identified for testing based on review of triage notations, electronic medical records, or

Table 1. Continued

	Diagnosis of HIV						Missing Data	
	Yes (n = 504)			No (n = 92,131)			No.	%
	No.	%	Median (IQR)	No.	%	Median (IQR)		
Previous sexually transmitted infections								
Gonorrhea	160	32		18,352	20		0	0
Chlamydia	69	14		20,282	22		0	0
Herpes simplex	35	7		5,077	6		0	0
Syphilis	35	7		1,801	2		0	0
Genital warts	48	10		8,989	10		0	0
Symptoms								
Anal discharge	0	0		12	0		0	0
Anal lesion	0	0		9	0		37	0.04
Dysuria	96	19		15,417	17		0	0
Genital pruritus	63	13		14,491	16		0	0
Genital rash	44	9		4,237	5		0	0
Urethral discharge	124	25		24,745	27		0	0
Any genitourinary symptoms	208	41		39,489	43		37	
Other risk factors								
Injection drug use	59	12		5,035	6		95	0.1
Prostitution	41	8		2,822	3		93	0.1
Sex with prostitute	41	8		7,208	8		109	0.1
Sex with HIV-infected partner	111	22		1,931	2		112	0.1
Sex with injection drug-using partner	56	11		6,200	7		124	0.1
Past HIV testing	391	78		61,856	67		30	0.03

Abbreviations: HIV, human immunodeficiency virus; IQR, interquartile range.

^a American Indian or Alaska Native, Native Hawaiian, or non-Hawaiian Pacific Islander.

referral by emergency department staff. The primary means of selection include identification of risk characteristics, clinician concern, and patient request. Patients are also referred when identified as having signs or symptoms suggestive of HIV infection. Any of a broad list of criteria may prompt testing, but patients are not systematically assessed for all possible risk indicators. Those who consent undergo structured and comprehensive risk assessment in conjunction with prevention counseling. The undiagnosed HIV infection prevalence in this setting is approximately 0.7%.

The risk score was applied to each patient included in this validation sample. Patients were then grouped into the unique risk categories identified during model development, and the proportions within each group are reported as percentages with 95% confidence intervals. Similar to our approach to internal validation, calibration was assessed by graphically comparing predicted HIV prevalence with observed HIV prevalence, by generating a linear regression line, and by calculating its slope and R^2 . Similarly, discrimination was assessed by constructing an ROC curve and calculating the area under the curve.

Data management and statistical analyses. All data were managed using Microsoft Access (Microsoft Corporation, Redmond, Washington), and statistical analyses were performed using SAS, version 9.2 (SAS Institute, Inc., Cary,

North Carolina), Stata, version 10.1 (Stata Corporation, College Station, Texas), or SPSS, version 18 (SPSS Inc., Chicago, Illinois). Fractional polynomials using the “fracpoly” command in Stata were used to model continuous and ordinal variables (e.g., age and lifetime number of sexual partners) and the outcome (25, 31). Because the extent of missing data was small and likely not influential relative to the estimates from the regression analysis, we used a Markov chain Monte Carlo approach to multiple imputation as a means to provide unbiased estimates of regression coefficients in the final model and to confirm their stability between the complete-case analysis and the analysis where imputation was used (32, 33). In addition, an unconditional bootstrapping approach was used to estimate 95% confidence intervals for the regression coefficients of the final model (34). This study was approved by the institutional review boards from each institution.

RESULTS

The derivation sample consisted of 92,635 patients. Of these, 504 (0.54%) were diagnosed with HIV infection during the clinical encounter. The median age was 27 years (interquartile range, 22–35; range, 13–87), and 62% were male. In terms of race/ethnicity, 42% were white, 29% were

Table 2. Multivariate Logistic Regression Model for the Prediction of Newly Diagnosed Human Immunodeficiency Virus Infection and Translation of Regression Coefficients into the Denver HIV Risk Score, Denver, Colorado, 1996–2008

Variable	β	95% Confidence Interval	Score
Age, years			
<22 or >60	0	Referent	0
22–25 or 55–60	0.4	0.3, 0.8	+4
26–32 or 47–54	1.0	0.7, 1.3	+10
33–46	1.2	0.9, 1.5	+12
Gender			
Female	0	Referent	0
Male	2.1	1.8, 2.4	+21
Race/ethnicity			
Black	0.9	0.7, 1.1	+9
Hispanic	0.3	0.1, 0.5	+3
Other ^a	-0.1	0.3, 0.1	0
White	0	Referent	0
Sexual practices			
Sex with a male	2.2	2.0, 2.5	+22
Vaginal intercourse	-1.0	0.8, -1.2	-10
Receptive anal intercourse	0.8	0.6, 1.0	+8
Other risk factors			
Injection drug use	0.9	0.7, 1.1	+9
Past HIV testing	-0.4	0.2, -0.6	-4

Abbreviation: HIV, human immunodeficiency virus.

^a American Indian or Alaska Native, Native Hawaiian, or non-Hawaiian Pacific Islander.

Hispanic, 25% were black, and 4% represented another race or ethnicity.

Table 1 shows the distribution of all candidate predictor variables by HIV serostatus. Patients infected with HIV were generally older and more likely to be male, to be white, to have had a larger number of sexual partners, to have had sex with males, and to have participated in insertive or receptive oral or anal intercourse. Patients with HIV infection were also more likely to have injected drugs, been tested previously for HIV infection, used condoms, had a previous diagnosis of gonorrhea or syphilis, and served as a prostitute or had sex with an HIV-infected partner or someone who injected drugs. Genital rash was the only symptom that was more common among patients infected with HIV.

The following variables were independently associated with HIV diagnosis and were included in the full risk model: age, gender, race/ethnicity, sex with a male, sex with a female in the previous year, sex with an HIV-infected partner, vaginal intercourse, insertive anal intercourse, receptive anal intercourse, oral intercourse, injection drug use, prostitution in the previous year, prior HIV testing, and history of syphilis, genital rash, or genital discharge. The slope of this model was 0.82, its R^2 was 0.88, and the area under the ROC curve was 0.86 (95% confidence interval (CI): 0.84, 0.88).

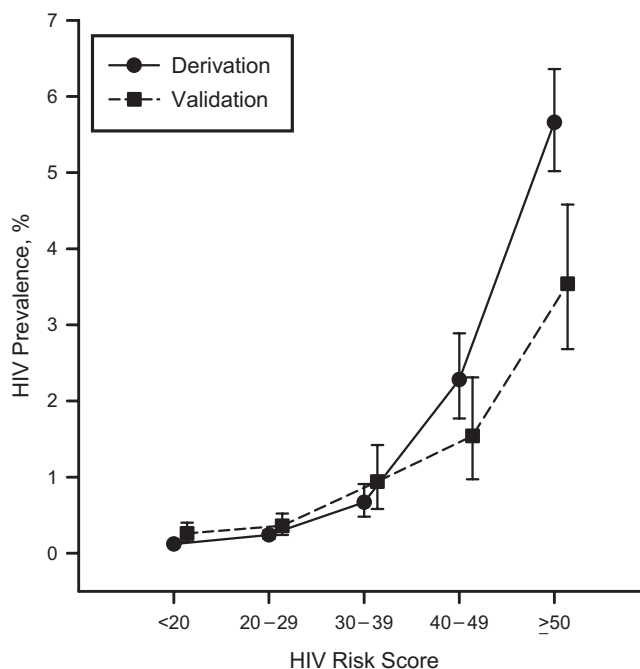
**Figure 1.** Prevalence of human immunodeficiency virus (HIV) infection within each risk score category in the derivation and validation samples, Denver, Colorado, 1996–2008 and Cincinnati, Ohio, 1998–2010. Point estimates and 95% confidence intervals are staggered for clarity. The risk score ranges from -14 to +81. Bars, 95% confidence interval.

Table 2 shows the results of the final pruned risk model. This model included age (categorized as <22 or >60 years, 22–25 or 55–60 years, 26–32 or 47–54 years, or 33–46 years); gender (male or female); race/ethnicity (black, Hispanic, white, or other); sexual practices (sex with a male, receptive anal intercourse, and vaginal intercourse); and other risk behaviors (injection drug use or past HIV testing), with a composite score ranging from -14 to +81. This model demonstrated outstanding calibration (calibration regression slope of 0.95 and an R^2 of 0.94) and discrimination (area under the ROC curve = 0.85, 95% CI: 0.83, 0.88).

All patients were then assigned scores according to the risk model and categorized into the following 5 mutually exclusive groups: <20 (very low risk); 20–29 (low risk); 30–39 (moderate risk); 40–49 (high risk); and ≥ 50 (very high risk). The HIV prevalences within these groups were 0.12% (95% CI: 0.08, 0.15) ($n = 54/46,627$), 0.24% (95% CI: 0.19, 0.30) ($n = 78/32,446$), 0.67% (95% CI: 0.48, 0.91) ($n = 40/5,965$), 2.28% (95% CI: 1.77, 2.89) ($n = 66/2,897$), and 5.66% (95% CI: 5.02, 6.36) ($n = 266/4,700$), respectively (Figure 1).

The top 3 risk groups (i.e., patients scoring 30 points or greater) represented 73.8% (95% CI: 69.7, 77.6) ($n = 372/504$) of all patients diagnosed with HIV infection yet only 14.6% (95% CI: 14.4, 14.9%) ($n = 13,562/92,635$) of the total derivation sample; the top 2 risk groups (i.e., patients scoring 40 points or greater) represented 65.9% (95% CI: 61.6, 70.0) ($n = 332/504$) of those diagnosed with HIV

Table 3. Risk Score Variables for the Derivation and Validation Cohorts According to Diagnosis of Human Immunodeficiency Virus Infection, Denver, Colorado, 1996–2008

	Derivation Cohort				Validation Cohort			
	HIV-Infected (n = 504)		Not HIV-Infected (n = 92,131)		HIV-Infected (n = 168)		Not HIV-Infected (n = 22,815)	
	No.	% ^a	No.	%	No.	%	No.	%
Age group, years								
<22 or >60	45	8.9	22,464	24.4	31	18.5	5,790	25.4
22–25 or 55–60	72	14.3	21,072	22.9	34	20.2	4,840	21.2
26–32 or 47–54	184	36.5	26,513	28.8	43	25.6	6,597	28.9
33–46	203	40.3	22,082	24.0	60	35.7	5,588	24.5
Gender								
Female	74	14.7	35,611	38.7	33	19.6	11,196	49.1
Male	430	85.3	56,520	61.4	135	80.4	11,619	50.9
Race/ethnicity								
Black	119	23.6	22,620	24.6	86	51.2	11,942	52.3
Hispanic	129	25.6	26,652	28.9	3	1.8	333	1.5
Other ^b or white	256	50.8	42,859	46.5	79	47.0	10,540	46.2
Sexual practices								
Sex with a male	404	80.6	42,721	46.4	111	66.1	11,672	51.2
Vaginal intercourse	178	35.3	80,334	87.2	117	69.6	18,519	81.2
Receptive anal intercourse	279	55.4	12,854	14.0	68	40.5	3,402	14.9
Other risk factors								
Injection drug use	59	11.8	5,035	5.5	10	6.0	1,602	7.0
Past HIV testing	391	77.6	61,856	67.2	105	62.5	16,212	71.1

Abbreviation: HIV, human immunodeficiency virus.

^a Percentages may not add to 100% because of rounding.

^b American Indian or Alaska Native, Native Hawaiian, or non-Hawaiian Pacific Islander.

infection yet only 8.2% (95% CI: 8.0, 8.4) ($n = 7,597/92,635$) of the sample; and the top risk group (i.e., patients scoring 50 points or greater) represented 52.7% (95% CI: 48.3, 57.2) ($n = 266/504$) of those diagnosed with HIV infection yet only 5.1% (95% CI: 4.9, 5.2%) ($n = 4,700/92,635$) of the sample.

The validation sample consisted of 22,983 patients, of whom 168 (0.73%) were identified with HIV infection. When categorized into the same 5 risk groups, HIV prevalence was 0.31% (95% CI: 0.20, 0.45) ($n = 27/8,782$) for persons with a score of <20, 0.41% (95% CI: 0.29, 0.57) ($n = 36/8,677$) for those with a score of 20–29, 0.99% (95% CI: 0.63, 1.47) ($n = 24/2,431$) for those with a score of 30–39, 1.59% (95% CI: 1.02, 2.36) ($n = 24/1,505$) for those with a score of 40–49, and 3.59% (95% CI: 2.73, 4.63) ($n = 57/1,588$) for those with a score of ≥ 50 (Figure 1).

In this sample, the top 3 risk groups represented 62.5% (95% CI: 54.7, 69.8) ($n = 105/168$) of all patients diagnosed with HIV infection yet only 24.0% (95% CI: 23.5, 24.6) ($n = 5,524/22,983$) of the sample; the top 2 risk groups represented 48.2% (95% CI: 40.5, 56.0) ($n = 81/168$) of those diagnosed with HIV infection yet only 13.5% (95% CI: 13.0, 13.9) ($n = 3,093/22,815$) of the sample; and the top group represented 33.9% (95% CI: 26.8, 41.6) ($n = 57/168$) of those

diagnosed with HIV infection yet only 6.9% (95% CI: 6.6, 7.2) ($n = 1,588/22,815$) of the sample.

Risk score variables for both the derivation sample and the validation sample are shown in Table 3. The calibration regression slope for the validation sample was 1.07 and its R^2 was 0.98 (Figure 2A). The area under the ROC curve for the validation sample was 0.75 (95% CI: 0.70, 0.78) (Figure 2B).

DISCUSSION

To our knowledge, this is the first study to empirically derive and externally validate an HIV risk prediction tool for estimating the prevalence of undiagnosed HIV infection among patients seeking medical care. Prior work has been conducted to evaluate risk and to improve targeting of certain patients, but such work has focused either on specific clinical settings or on specific patient populations, and to our knowledge such practices have not been further evaluated or widely used (16, 19). Our risk score categorizes patients, based on a limited number of variables, into groups with increasing HIV prevalence, and its use has the potential to improve the effectiveness and efficiency of HIV screening across settings where HIV testing is performed by accurately identifying patients at risk and helping to ensure that they are targeted in

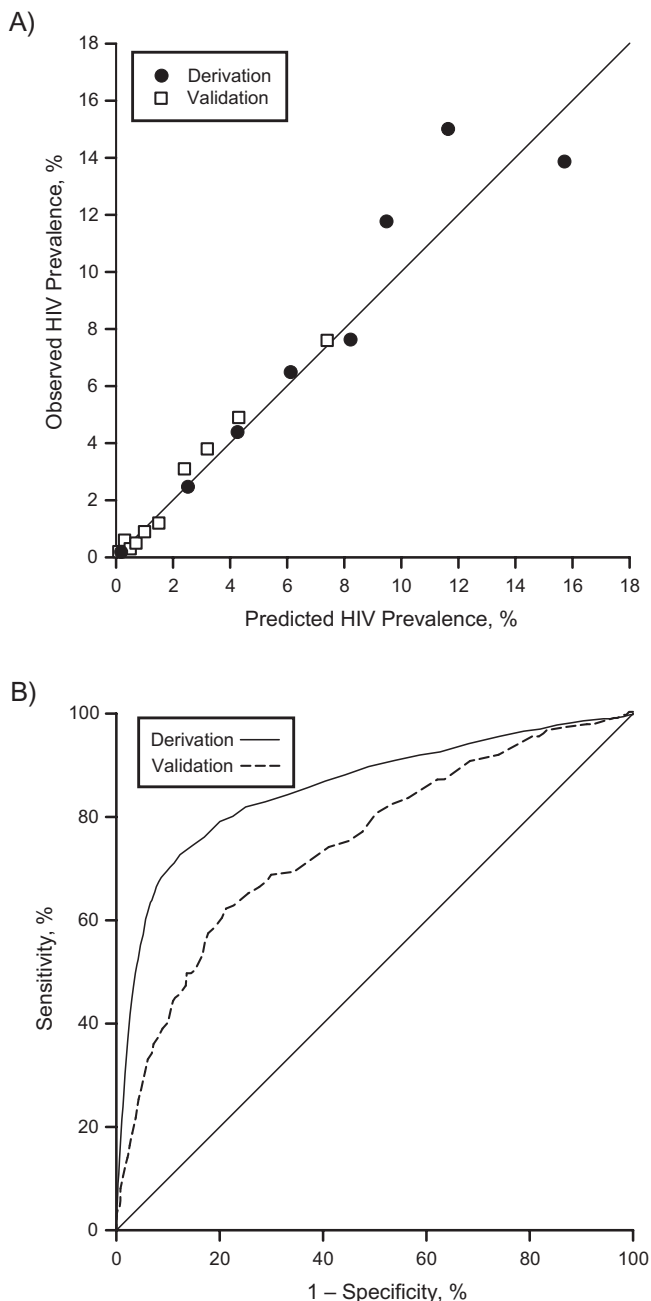


Figure 2. Calibration (part A) and discrimination (part B) of a risk score designed to identify patients at risk of human immunodeficiency virus (HIV) infection, Denver, Colorado, 1996–2008 and Cincinnati, Ohio, 1998–2010. The area under the receiver operating characteristic curve was 0.85 (95% confidence interval: 0.83, 0.88) in the derivation sample and 0.75 (95% confidence interval: 0.70, 0.78) in the validation sample.

a systematic, high-yield fashion. This approach may serve best in nontraditional clinical care settings such as emergency departments, urgent care centers, or general practices where implementation of nontargeted screening is most challenging.

Current CDC recommendations for HIV testing call for nontargeted opt-out screening in most health-care settings

(1). Other advisory groups, including the World Health Organization and the US Preventive Services Task Force, have tempered their recommendations in support of targeted strategies (10, 11). While the ability to screen every patient for HIV infection would clearly yield the largest numbers of diagnoses and have the greatest impact on mitigating the epidemic, we believe abandoning targeted HIV screening is premature, especially in light of ongoing barriers related to implementation of nontargeted HIV screening (2). In addition, prior work by our team and by others evaluating nontargeted HIV screening in real-world clinical settings demonstrates that this method is costly and probably fails to identify a large number of persons with HIV infection (7, 9, 35, 36).

Traditional targeted screening methods are thought to be ineffective, and arguments to abandon this approach remain active because it is thought that clinicians are too busy to perform risk assessments and patients are unwilling or unable to provide accurate risk information. However, previous research indicates that most patients are willing to discuss and disclose risk behaviors when asked (18, 37, 38), while novel approaches for screening, including use of computerized kiosks and health educators (39–41), have already improved the fidelity of assessing HIV risk. Additionally, traditional risk assessments have relied on subjective assessment of risk by clinicians and have not been systematically applied, something the risk score may obviate.

The risk score may also help prioritize HIV testing and prevention resources in the United States and other countries where the epidemic is concentrated, and where clinicians and public health leaders must make important decisions about where to focus HIV testing efforts. The risk score may be administered using paper forms, although integration into an electronic medical care system will likely be optimal. The latter approach would allow for real-time calculation of a patient's score and provide the opportunity to limit the number of questions asked, depending on initial responses to questions.

Although the fundamental premise of the risk score is to identify those most at risk, it may also be used to identify patients at low risk, thus limiting the number of tests performed where the yield of new diagnoses is low; this approach also would reduce the likelihood of false-positive tests encountered during large-scale screening (36), especially when performed in low-prevalence settings. Furthermore, the risk score may have the additive benefit of informing clinicians and patients about the risk and allow both to share decision-making when considering testing for HIV infection or when establishing a risk reduction plan. Decision and cost-effectiveness modeling and application of the risk score in clinical practice will help define its most optimal use, which is likely to vary depending on the setting and its underlying HIV prevalence (6).

The risk score includes only 8 variables and was shown to have excellent test characteristics. The model included 3 demographic variables and 5 risk behavior variables and reflects national demographic and risk behavior estimates from the CDC. The estimated numbers of HIV and acquired immunodeficiency syndrome cases are highest among persons aged 20–49 years, with the peak number of cases occurring between ages 35 and 44 years (42). We used a nonlinear

approach to model the relation between age and HIV infection, with findings consistent with those reported nationally. Additionally, the incidence of HIV infection is estimated to be highest among men and those who are black or Hispanic, all findings confirmed by our model (42). Finally, risky behaviors, including sex with a male (regardless of whether the person being tested was male or female), anal intercourse, and injection drug use, were also associated with HIV infection, whereas vaginal intercourse and past HIV testing were negatively associated with HIV infection.

The risk score includes patient demographic characteristics, which are relatively easy to ascertain and are generally collected as part of standard practice where HIV testing is performed, but it also includes 5 potentially sensitive risk-based characteristics that may be variably disclosed by patients. In particular, variable disclosure of “sex with a male” if the patient is male, “receptive anal intercourse,” and “injection drug use” may limit the acceptability of routinely implementing this instrument. Prior to including risk-based characteristics in the model, we developed a demographic-only model, which did not perform well (data not shown). Our results support the need to ascertain and highlight the importance of ascertaining at least some behavioral characteristics to estimate a patient’s risk of HIV infection.

Although the risk score was developed and tested using 2 large and distinct patient samples, its predictive accuracy also differed slightly and thus may differ in other communities or clinical settings. In addition, because HIV testing was optional in both clinical settings, selection bias may have been introduced. The generalizability of the risk score is supported, however, by the fact that the model’s calibration between the two study groups was similar. Because this study was not powered to discriminate statistically between the 5 risk groups defined during derivation and because there was a slight loss in discrimination between the derivation and validation cohorts, additional validation may be warranted. Future investigation may require the model’s refinement prior to clinical implementation. Although it is believed that risk for acquiring HIV infection may change over time, the face validity of the risk score in relation to the epidemiology of HIV infection in the United States suggests that the shift in risk is relatively slow. Regardless, prior to or during the initial phases of clinical implementation, the ability of the model to risk-stratify patients should be assessed and adjusted as necessary, including recalibration to maximize the predictive accuracy of the score in each unique setting.

In summary, we derived and validated a risk score that categorized patients into groups with increasing probabilities of HIV infection. The risk score may help clinicians identify patients for directed HIV testing or help public health leaders focus limited resources to improve and streamline approaches to HIV screening.

ACKNOWLEDGMENTS

Author affiliations: Department of Emergency Medicine, Denver Health Medical Center, Denver, Colorado (Jason S. Haukoos, Emily Hopkins, Brooke Bender, Lynsay A.

MacLaren, Richard L. Byyny); Department of Emergency Medicine, School of Medicine, University of Colorado, Aurora, Colorado (Jason S. Haukoos, Comilla Sasson, Richard L. Byyny); Department of Epidemiology, Colorado School of Public Health, Aurora, Colorado (Jason S. Haukoos); Department of Emergency Medicine, University of Cincinnati Medical Center, Cincinnati, Ohio (Michael S. Lyons, Christopher J. Lindsell); Department of Emergency Medicine, Johns Hopkins Hospital, Baltimore, Maryland (Richard E. Rothman, Yu-Hsiang Hsieh); Denver Public Health, Denver Health Medical Center, Denver, Colorado (Mark W. Thrun); and Department of Medicine, School of Medicine, University of Colorado, Aurora, Colorado (Mark W. Thrun).

This research was supported, in part, by an Independent Scientist Award (K02HS017526) from the Agency for Healthcare Research and Quality to Dr. Haukoos.

Conflict of interest: none declared.

REFERENCES

1. Branson BM, Handsfield HH, Lampe MA, et al. Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. *MMWR Recomm Rep*. 2006;55(RR-14):1–17.
2. Bartlett JG, Branson BM, Fenton K, et al. Opt-out testing for human immunodeficiency virus in the United States: progress and challenges. *JAMA*. 2008;300(8):945–951.
3. Haukoos JS, Hopkins E, Hull A, et al. HIV testing in emergency departments in the United States: a national survey. *Ann Emerg Med*. 2011;58(1 suppl 1):S10–S16.
4. Rothman RE, Hsieh YH, Harvey L, et al. 2009 US emergency department HIV testing practices. *Ann Emerg Med*. 2011; 58(1 suppl 1):S3–S9.
5. Bogart LM, Howerton D, Lange J, et al. Scope of rapid HIV testing in urban U.S. hospitals. *Public Health Rep*. 2008;123(4): 494–503.
6. Haukoos JS, Hopkins E, Byyny RL, et al. Design and implementation of a controlled clinical trial to evaluate the effectiveness and efficiency of routine opt-out rapid human immunodeficiency virus screening in the emergency department. *Acad Emerg Med*. 2009;16(8):800–808.
7. Haukoos JS, Hopkins E, Conroy AA, et al. Routine opt-out rapid HIV screening and detection of HIV infection in emergency department patients. *JAMA*. 2010;304(3):284–292.
8. Haukoos J, Conroy A, Hopkins E, et al. Cost effectiveness of routine opt-out rapid HIV screening in the emergency department: results from a prospective controlled clinical trial [abstract]. *Acad Emerg Med*. 2009;16(suppl 1):S147–S148.
9. Brown J, Shesser R, Simon G, et al. Routine HIV screening in the emergency department using the new US Centers for Disease Control and Prevention Guidelines: results from a high-prevalence area. *J Acquir Immune Defic Syndr*. 2007; 46(4):395–401.
10. World Health Organization. *Guidance on Provider-Initiated HIV Testing and Counselling in Health Facilities*. Geneva, Switzerland: World Health Organization; 2007.
11. Chou R, Huffman L. *Screening for Human Immunodeficiency Virus: Focused Update of a 2005 Systematic Review for the U.S. Preventive Services Task Force*. Rockville, MA: Agency for Healthcare Research and Quality; 2007.

12. Joint United Nations Programme on HIV/AIDS and World Health Organization. *2009 AIDS Epidemic Update*. Geneva, Switzerland: World Health Organization; 2009. (http://www.unaids.org/en/media/unaids/contentassets/dataimport/pub/report/2009/jc1700_epi_update_2009_en.pdf). (Accessed March 1, 2011).
13. Additional recommendations to reduce sexual and drug abuse-related transmission of human T-lymphotropic virus type III/lymphadenopathy-associated virus. *MMWR Morb Mortal Wkly Rep*. 1986;35(10):152–155.
14. Recommendations for HIV testing services for inpatients and outpatients in acute-care hospital settings. *MMWR Recomm Rep*. 1993;42(RR-2):1–6.
15. Revised guidelines for HIV counseling, testing, and referral. *MMWR Recomm Rep*. 2001;50(RR-19):1–57.
16. Chen Z, Branson B, Ballenger A, et al. Risk assessment to improve targeting of HIV counseling and testing services for STD clinic patients. *Sex Transm Dis*. 1998;25(10):539–543.
17. Peterman TA, Todd KA, Mupanduki I. Opportunities for targeting publicly funded human immunodeficiency virus counseling and testing. *J Acquir Immune Defic Syndr Hum Retrovirol*. 1996;12(1):69–74.
18. Gerbert B, Bronstone A, McPhee S, et al. Development and testing of an HIV-risk screening instrument for use in health care settings. *Am J Prev Med*. 1998;15(2):103–113.
19. Menza TW, Hughes JP, Celum CL, et al. Prediction of HIV acquisition among men who have sex with men. *Sex Transm Dis*. 2009;36(9):547–555.
20. Wasson JH, Sox HC, Neff RK, et al. Clinical prediction rules. Applications and methodological standards. *N Engl J Med*. 1985;313(13):793–799.
21. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA*. 1997;277(6):488–494.
22. Janssen KJ, Moons KG, Harrell FE Jr. Prediction rules must be developed according to methodological guidelines [letter]. *Ann Intern Med*. 2010;152(4):263.
23. Greenland S. Modeling and variable selection in epidemiologic analysis. *Am J Public Health*. 1989;79(3):340–349.
24. Austin PC, Tu JV. Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *J Clin Epidemiol*. 2004;57(11):1138–1146.
25. Royston P, Sauerbrei W, Altman DG. Modeling the effects of continuous risk factors. *J Clin Epidemiol*. 2000;53(2):219–221.
26. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006;25(1):127–141.
27. Steyerberg E. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York, NY: Springer Publishing Company; 2009.
28. Harrell F. Regression coefficients and scoring rules [letter]. *J Clin Epidemiol*. 1996;49(7):819.
29. Hosmer D, Lemeshow S. *Applied Logistic Regression*. 2nd ed. New York, NY: John Wiley & Sons, Inc; 2000.
30. Lyons MS, Lindsell CJ, Ledyard HK, et al. Emergency department HIV testing and counseling: an ongoing experience in a low-prevalence area. *Ann Emerg Med*. 2005;46(1):22–28.
31. Royston P, Sauerbrei W. Building multivariable regression models with continuous covariates in clinical epidemiology—with an emphasis on fractional polynomials. *Methods Inf Med*. 2005;44(4):561–571.
32. Haukoos JS, Newgard CD. Advanced statistics: missing data in clinical research—part 1: an introduction and conceptual framework. *Acad Emerg Med*. 2007;14(7):662–668.
33. Newgard CD, Haukoos JS. Advanced statistics: missing data in clinical research—part 2: multiple imputation. *Acad Emerg Med*. 2007;14(7):669–678.
34. Haukoos JS, Lewis RJ. Advanced statistics: bootstrapping confidence intervals for statistics with “difficult” distributions. *Acad Emerg Med*. 2005;12(4):360–365.
35. White DA, Scribner AN, Vahidnia F, et al. HIV screening in an urban emergency department: comparison of screening using an opt-in versus an opt-out approach. *Ann Emerg Med*. 2011;58(1 suppl 1):S89–S95.
36. Walensky RP, Arbelaez C, Reichmann WM, et al. Revising expectations from rapid HIV tests in the emergency department. *Ann Intern Med*. 2008;149(3):153–160.
37. Gerbert B, Brown B, Volberding P, et al. Physicians’ transmission prevention assessment and counseling practices with their HIV positive patients. *AIDS Educ Prev*. 1999;11(4):307–320.
38. Gerbert B, Maguire BT, Coates TJ. Are patients talking to their physicians about AIDS? *Am J Public Health*. 1990;80(4):467–469.
39. Hopkins E, Bender B, Haukoos J. Evaluation of kiosk-based opt-out and opt-in rapid HIV screening for unscheduled ambulatory care [abstract]. Presented at the 2010 National Summit on HIV Diagnosis, Prevention, and Access to Care, National Harbor, Maryland, November 18, 2010.
40. Merchant RC, Clark MA, Langan TJ IV, et al. Effectiveness of increasing emergency department patients’ self-perceived risk for being human immunodeficiency virus (HIV) infected through audio computer self-interview-based feedback about reported HIV risk behaviors. *Acad Emerg Med*. 2009;16(11):1143–1155.
41. Calderon Y, Haughey M, Bijur PE, et al. An educational HIV pretest counseling video program for off-hours testing in the emergency department. *Ann Emerg Med*. 2006;48(1):21–27.
42. Centers for Disease Control and Prevention. *Table 1a. Diagnoses of HIV Infection, by Year of Diagnosis and Selected Characteristics, 2005–2008—37 States With Confidential Name-based HIV Infection Reporting*. Atlanta, GA: Centers for Disease Control and Prevention; 2008. (<http://www.cdc.gov/hiv/surveillance/resources/reports/2008report/table1a.htm>). (Accessed October 15, 2010).