EBioMedicine 68 (2021) 103408

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

Predictive nomogram for leprosy using genetic and epidemiological risk factors in Southwestern China: Case–control and prospective analyses

Si-Yu Long[a], Ji-Ya Sun[d], Le Wang[a,b], Heng Long[e], Hai-Qin Jiang[a], Ying Shi[a], Wen-Yue Zhang[a], Jing-Shu Xiong[a], Pei-Wen Sun[a,b], Yan-Qing Chen[a], You-Ming Mei[a], Chun Pan[a], Zhen-Zhen Wang[a], Zi-Wei Wu[a], Ai-Ping Wu[d], Mei-Wen Yu[a,b,*], Hong-Sheng Wang[a,b,c,*]

[a] *Jiangsu Key Laboratory of Molecular Biology for Skin Diseases and STIs, Institute of Dermatology, Chinese Academy of Medical Sciences & Peking Union Medical College, Nanjing, 210042, China.*
[b] *National Centre for Leprosy Control, China CDC, Nanjing, China*
[c] *Centre for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China*
[d] *Center for Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College; Suzhou Institute of Systems Medicine, Suzhou, China*
[e] *Wenshan Institute of Dermatology, Wenshan, Yunnan, China*

ARTICLE INFO

Article History:
Received 18 January 2021
Revised 3 May 2021
Accepted 6 May 2021
Available online xxx

Keywords:
Leprosy
Nomogram
Risk stratification
Leprosy prediction

ABSTRACT

Background: There is a high incidence of leprosy among house-contacts compared with the general population. We aimed to establish a predictive model using these genetic factors along with epidemiological factors to predict leprosy risk of leprosy household contacts (HHCs).

Methods: Weighted genetic risk score (wGRS) encompassing genome wide association studies (GWAS) variants and five non-genetic factors were examined in a case–control design associated with leprosy risk including 589 cases and 647 controls from leprosy HHCs. We constructed a risk prediction nomogram and evaluated its performance by concordance index (C-index) and calibration curve. The results were validated using bootstrap resampling with 1000 resamples and a prospective design including 1100 HHCs of leprosy patients.

Finding: The C-index for the risk model was 0·792 (95% confidence interval [CI] 0·768-0·817), and was confirmed to be 0·780 through bootstrapping validation. The calibration curve for the probability of leprosy showed good agreement between the prediction of the nomogram and actual observation. HHCs were then divided into the low-risk group (nomogram score ≤ 81) and the high-risk group (nomogram score > 81). In prospective analysis, 12 of 1100 participants had leprosy during 63 months' follow-up. We generated the nomogram for leprosy in the validation cohort (C-index 0·773 [95%CI 0·658-0·888], sensitivity75·0%, specificity 66·8%). Interpretation The nomogram achieved an effective prediction of leprosy in HHCs. Using the model, the risk of an individual contact developing leprosy can be determined, which can lead to a rational preventive choice for tracing higher-risk leprosy contacts.

Funding: The ministry of health of China, ministry of science and technology of China, Chinese academy of medical sciences, Jiangsu provincial department of science and technology, Nanjing municipal science and technology bureau.

1. Introduction

Leprosy, which is caused mainly by Mycobacterium leprae, can progress to peripheral nerve injury and systematic deformity in untreated individuals [1]. According to the official WHO records, a total of 208641 new leprosy patients were reported globally in 2018, China contributed 521 (0·25%) of these cases [2]. Leprosy is mainly prevalent in southwestern provinces of China, including Sichuan, Hunan, Yunnan, and Guizhou [3]. Current evidence suggests that the leprosy household contacts (HHCs) have a higher-risk of developing the disease than the general population [4]. Therefore, contact tracing and post-exposure prophylaxis (PEP) serve as the basis for leprosy control. A distinct risk prediction model that could identify higher-risk contacts needs to be developed, which would allow physicians to

* Corresponding authors at: Chinese Academy of Medical Sciences and Peking Union Medical College, Institute of Dermatology, St 12 Jiangwangmiao, Nanjing, Jiangsu, 210042, China.
E-mail addresses: yumw@pumcderm.cams.cn (M.-W. Yu), whs33@vip.sina.com (H.-S. Wang).

https://doi.org/10.1016/j.ebiom.2021.103408

Research in context

Evidence before this study

We searched PubMed and SpringerLink for studies for abstracts and articles using the search terms “leprosy”, “contact”, “epidemiology”, “SNP”, “nomogram” and “risk factors”, from database inception to Aug 1, 2020, with no language restrictions. We found no nomogram-based studies focusing on predicting for leprosy developing in leprosy contacts. Most articles focused on the risk factors associated with leprosy among household contacts of patients. To date, risk factors that have been proposed for acquisition of leprosy include type of leprosy in index patient, age, sex, genetic influences, socioeconomic conditions and some serologic markers. A modeling study attempted to predict leprosy in the Chinese population based on a weighted genetic risk score, but the study had no information of individual's exposure to *M. leprae* and the outcomes could not truly examine the discriminatory power of predicting the incidence of leprosy. Although many groups have tried to estimate the risk factors of leprosy developing in leprosy household contacts, none has specifically constructed a model to predict leprosy contacts at risk of developing leprosy.

Added value of this study

Our study is the first to establish a predictive nomogram for leprosy based on the data of leprosy contact individuals using genetic and epidemiological risk factors and determine the accuracy of the prediction model in a prospective study.

Implications of all the available evidence

The findings of this study show that the individuals at high risk of developing leprosy are more likely to be younger, male, minority and higher wGRS, their index cases had higher baseline delay period of detection and likely to be lepromatous leprosy patients. We established a nomogram based on multivariate logistic regression analysis. The nomogram is valuable for risk stratification management, and will be helpful for controlling leprosy transmission by formulating strategies of tracing contacts with the high risk of developing leprosy.

determine the PEP target group and formulate strategies for tracing contacts.

As an infectious disease, an individual's exposure to *M. leprae* and epidemiological factors play important roles in the development of leprosy. Studies showed that the increased incidence of leprosy in contact individuals is likely associated with leprosy classification of index cases [4]. Gender differences and age also play important roles in the risk of developing leprosy among contacts [4–7]. Genome wide association studies (GWAS) have identified a number of single nucleotide polymorphisms (SNPs) associated with the genetic predisposition to leprosy. [8–14] Based on these findings, we calculated the weighted genetic risk score (wGRS) with the published GWAS results and combined non-genetic risk factors to construct a predictive nomogram model in a Chinese leprosy HHCs cohort. We assessed the predictive accuracy of this model and validated it in an independent follow-up group.

2. Methods

2.1. Study populations

A case–control study was conducted on a primary cohort of leprosy HHCs, enrolled from January 2010 to June 2014 in four provinces (Yunnan, Guizhou, Sichuan and Hunan). Fig. 1 shows the enrollment and outcomes in this study. 644 leprosy HHCs detected as new leprosy patients in January 2010 to June 2014 in the study area were selected as case group. 685 Controls without leprosy were selected from the HHCs whose index cases were registered in January 2010 to June 2014 in the area. Leprosy cases and control individuals in the training cohort were matched according to their socioeconomic status and environmental conditions. All subjects were followed up from recruitment to December 31, 2019. Patients were diagnosed with leprosy by initial clinical evaluation based on clinical manifestations, slit skin smears and histopathological examinations. HHCs were defined as people living under the same roof and sharing food with the patient for at least six months among the past six years. Exclusions were those who refused to provide informed consent and any person that received treatment for tuberculosis or leprosy within one year. The controls ever diagnosed with leprosy until December 31, 2019 were excluded.

An independent cohort included 1164 leprosy HHCs whose index cases were diagnosed in January 2010 to June 2014 and who were followed up from September 2014 to the time of detection of a subsequent case or until December 31, 2019 was prospectively studied, using the same risk factors, inclusion and exclusion criteria as described previously.

Blood samples were collected from all participating subjects after obtaining an informed written consent. All participating individuals provided personal data including sex, contact age (when index cases were diagnosed) and ethnicity. Data such as index cases' delay period of detection and Ridley-Jopling classification [15] of index cases was collected from medical record.

2.2. Ethics statement

This study was approved by the institutional ethical committee of the Institute of Dermatology, Chinese Academy of Medical Sciences and Peking Union Medical College, Yunnan Center for Disease Control and Prevention, Guizhou Center for Disease Control and Prevention, Hunan Center for Disease Control and Prevention, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital. (2014-KY-003). This study was conducted according to the Declaration of Helsinki. All adult participants provided written informed consent. Parents or guardians provided written informed consent on behalf of children who participated in the study.

2.3. SNP selection, genotyping, and quality control

We selected 17 SNPs from previously published GWAS studies and one study combined whole-exome sequencing and targeted next-generation sequencing within the GWAS loci [8–14], wherein a genome-wide significant association ($p < 5 \times 10^{-8}$) between the SNPs in the genes *RAB32, HIF1A, BATF3, LACC1, CTSB, TNFSF15, CDH18, SLC29A3, DEC1, FLG, NOD2, IL18RAP/IL18R1, NCKIPSD, CARD9* and leprosy. Based on multiplex polymerase chain reaction (PCR) to precisely genotype SNPs with next generation sequencing. After the PCR amplification, the products genotyped according to the manufactures' protocol using the Illumina Hiseq X-10 platform. All the 17 SNPs were genotyped in the training cohort. In the validation cohort, the subjects were only genotyped for the SNPs at significance level in the training cohort. Variants went through the following quality control filters in the training cohort: Genotypes were manually curated with call rates above 97%, and minor allele frequency (MAF) > 1%. Ultimately, two variants with MAF < 1% in the training cohort were eliminated (rs149308743 and rs145562243). Subjects with missing data on one or more genetic variants of interest were also excluded from the analysis. A total of 15 variants and 2336 subjects (589 participants with leprosy, 647 controls free of leprosy, and 1100 participants of follow-up group) were included in the analyses.

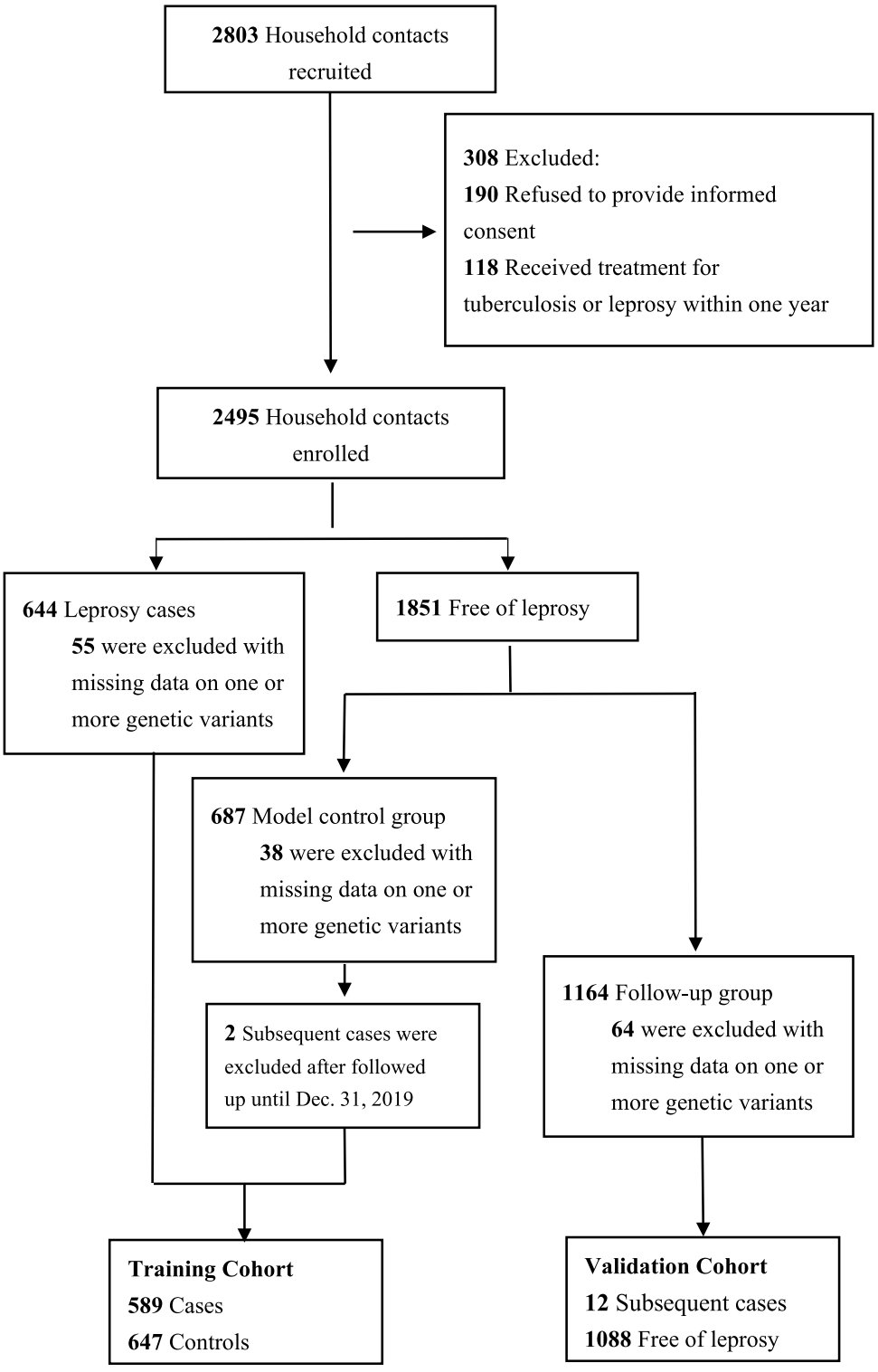

Fig. 1. Enrollment and outcomes. 63-month follow-up for all of the participants was performed including 5 clinical visits.

2.4. Statistical analysis

We tested the associations between phenotypes and SNPs using PLINK v 1·07 based on a logistic regression model adjusted for non-genetic factors, and constructed wGRS for each individual by summing the risk alleles (0/1/2) weighted by the β coefficient of each SNP with P value $<$ 0·05 obtained from the logistic regression in the training samples. The wGRS as a predictor for leprosy risk was analyzed.

Continuous variables including contact age, wGRS and index cases' delay period of detection were categorized into groups based on the epidemiological characteristics (contact age) or maximum Youden index (wGRS and index cases' delay period of detection). The least absolute shrinkage and selection operator (LASSO) method, was used to select the optimal predictive features in risk factors. The LASSO model used 10-fold cross-validation via minimum criteria. Features with nonzero coefficients in the LASSO regression model were selected [16]. Then, we did a binary multivariate logistic

regression analysis to test the independent significance of different factors and build a predicting model by incorporating the features including sex, contact age (when index cases were diagnosed), ethnicity, index cases' delay period of detection, Ridley-Jopling classification of index cases, and wGRS.

A nomogram model was formulated based on proportionally converting each regression coefficient in multivariate logistic regression to a 0-to100-point scale. The effect of the variable with the highest β coefficient (absolute value) is assigned 100 points. The points are added across independent variables to derive total points, which are converted to predicted probabilities [17]. The performance of nomogram was evaluated by discrimination (concordance index [C-index]) and calibration (calibration plots and Hosmer-Lemeshow calibration test) [18]. The discrimination of the nomogram was measured by the C-index and calibration with 1000 bootstrap samples to decrease the overfit bias. We also investigated the performance of every single factor, the model with all epidemiological risk factors and the relevant model including genetic and epidemiological risk factors. Calibration is useful for assessing whether actual outcomes approximate predicted outcomes for every nomogram [19]. In a well calibrated model, points are close to the 45-degree line. We find the optimal cutoff value to separate leprosy contacts into low-risk and high-risk groups as determined by the maximum sensitivity and specificity in the training cohort. During the independent validation of the nomogram, the total points of each HHC in the validation cohort were calculated according to the established nomogram. Then, logistic regression in this cohort was performed using the total points as a factor. Finally, the C-index and calibration curve were derived based on the regression analysis.

We used the "glmnet" package to perform the LASSO regression model analysis and "Hmisc" package to calculate C-index. Logistic regression, nomogram and calibration plots were made using the "rms" package of R software. We used SPSS to calculate Youden index and perform Hosmer-Lemeshow calibration test. 95% confidence intervals to the sensitivity and specificity were calculated by MedCalc software version 19·7·4. All statistical tests were conducted using R software version 3·6·1 and SPSS (version23; IBM Corporation). Statistical significance was set at 0·05.

2.5. Role of the funding source

The sponsors had no role in the study design, data collection, data analyses, interpretation, or writing of the study.

3. Results

1236 leprosy contacts met the inclusion criteria were included in the training cohort. For the validation cohort, we studied 1100 HHCs (Supplementary Fig.1). In the training cohort, the case group consisted of 589 leprosy patients, with a mean contact age (when the index cases were diagnosed) of 25·74 ± 16·067 years of age (range 0-80 years) while 647 contacts free of leprosy, with a mean age of 37·68 ± 15·919 years of age (range 7-82 years). In the validation cohort, the case group consisted of 12 leprosy patients, with a mean age of 25·83 ± 15·96 years of age (range 8-50 years) while 1088 contacts free of leprosy, with a mean age of 34·94 ± 16·11 years of age (range 1-77 years). The other baseline characteristics of participants are listed in Table 1.

The wGRS was calculated based on three SNPs which showed an association at $P < 0{\cdot}05$ in the training cohort after adjusting for sex, contact age (when index cases were diagnosed), ethnicity, index cases' delay period of detection, and Ridley-Jopling classification of index cases. These included rs2221593 at the *BATF3* locus (P = 0·03), rs146466242 at the *FLG* locus (P = 0·02) and rs9302752 at the *NOD2* locus ($P = 1{\cdot}54 \times 10^{-5}$). The characteristics and association results of the 15 variants are displayed in Table 2.

Table 1
Demographics and wGRS characteristics of leprosy contact[a].

Demographic Characteristics	No. (%) Training Cohort		Validation Cohort	
	Leprosy Cases (n = 589)	Controls (n = 647)	Subsequent leprosy cases (n = 12)	Free of leprosy (n = 1088)
Contact age(years)[b]				
≤10	116(19.7)	11(1.7)	4(33.3)	32(2.9)
11-20	137(23.3)	100(15.5)	1(8.3)	238(21.9)
21-30	122(20.7)	127(19.6)	3(25.0)	201(18.5)
31-40	101(17.1)	138(21.3)	0	185(17.0)
41-50	71(12.1)	105(16.2)	4(33.3)	222(20.4)
51-60	24(4.1)	108(16.7)	0	139(12.8)
61-70	15(2.5)	54(8.3)	0	66(6.1)
≥71	3(0.5)	4(0.6)	0	5(0.5)
Ethnicity				
Han	296(50.3)	366(56.6)	4(33.3)	613(56.3)
Minority	293(49.7)	281(43.4)	8(66.7)	475(43.7)
Sex				
Male	380(64.5)	317(49.0)	8(66.7)	569(52.3)
Female	209(35.5)	330(51.0)	4(33.3)	519(47.7)
Index cases' delay period of detection (months)				
≤32.5	304(51.6)	465(71.9)	11(91.7)	810(74.4)
>32.5	285(48.4)	182(28.1)	1(8.3)	278(25.6)
wGRS				
≤0.54	389(66.0)	513(79.3)	8(66.7)	809(74.4)
>0.54	200(34.0)	134(20.7)	4(33.3)	279(25.6)
Ridley-Jopling classification of index cases				
BB	27(4.6)	73(11.3)	1(8.3)	53(4.9)
BL	196(33.3)	211(32.6)	3(25.0)	370(34.0)
BT	39(6.6)	87(13.4)	1(8.3)	183(16.8)
I	6(1.0)	5(0.8)	0	4(0.4)
LL	276(46.9)	177(27.4)	7(58.3)	335(30.8)
TT	45(7.6)	94(14.5)	0	143(13.1)

[a] The table excluded individuals with missing data.
[b] Contact age when the index cases were diagnosed. OR, odds ratio; TT, Tuberculoid Leprosy; BT, Borderline tuberculoid Leprosy; BB, Mid-borderline Leprosy; BL, Borderline lepromatous Leprosy; LL, Lepromatous Leprosy; I, Indeterminate Leprosy; wGRS, Weighted genetic risk score.

Table 2
Association between SNPs and leprosy[a].

SNP	Region	Candidate gene	Risk allele	Minor allele	aa/Aa/AA[b]		F_A	F_U	OR (95%CI)[a]	*P*-value[a]
					Cases (n=589)	Controls (n=647)				
rs2275606	6q24.3	*RAB32*	A	A	64/229/296	53/244/350	0.30	0.27	1.11(0.91-1.34)	0.31
rs142179458	14q23.2	*HIF1A*	A	A	1/67/521	1/43/603	0.06	0.03	1.45(0.94-2.24)	0.10
rs2221593	1q32.3	*BATF3*	A	A	91/185/313	80/168/399	0.31	0.25	1.21(1.01-1.44)	0.03
rs3764147	13q14.11	*LACC1*	G	G	102/296/191	91/309/247	0.42	0.38	1.15 (0.96-1.38)	0.14
rs55894533	8p23.1	*CTSB*	C	C	138/293/158	147/324/176	0.48	0.48	1.10(0.92-1.32)	0.28
rs6478108	9q32	*TNFSF15*	C	T	124/246/219	143/308/196	0.42	0.46	1.15(0.97-1.37)	0.10
rs73058713	5p14.3	*CDH18*	C	A	12/142/435	16/143/488	0.14	0.14	1.00(0.78-1.29)	0.99
rs780668	10q22.1	*SLC29A3*	A	A	149/296/144	137/331/179	0.50	0.47	1.07(0.89-1.28)	0.50
rs10817758	9q32	*DEC1*	C	T	113/304/172	137/326/184	0.45	0.46	1.17(0.97-1.41)	0.09
rs146466242	1q21.3	*FLG*	A	A	3/23/563	0/10/637	0.02	0.01	2.70(1.17-6.23)	0.02
rs9302752	16q12.1	*NOD2*	C	C	63/243/283	33/236/378	0.31	0.23	1.57(1.28-1.92)	<0.001
rs2058660	2q12.1	*IL18RAP/IL18R1*	C	C	95/302/192	112/300/235	0.42	0.40	1.04(0.87-1.25)	0.67
rs4720118	7p14.3	*BBS9*	C	T	50/249/290	59/265/323	0.30	0.30	1.06(0.87-1.29)	0.57
rs663743	11q13.1	*CCDC88B*	A	A	31/211/347	36/205/406	0.23	0.21	1.10(0.89-1.36)	0.37
rs76418789	1p31.3	*IL23R*	A	A	2/55/532	1/54/592	0.05	0.04	1.01(0.66-1.54)	0.97

[a] Association tested for risk alleles with logistic regression model adjusted for sex, contact age, ethnicity, delay period of detection and Ridley-Jopling classification of index cases without the missing data in training cohort.
[b] aa/Aa/AA, the numbers of cases and controls with minor allele homozygote/heterozygote/major allele homozygote genotypes, respectively. F_A, minor allele frequency in 589 leprosy patients; F_U, minor allele frequency in 647 control subjects, OR, odds ratio; CI, confidence interval; SNP, single nucleotide polymorphism.

Six potential predictors based on the training cohort with nonzero coefficients were selected by the LASSO regression model (Supplementary Fig.2). The results of univariate logistic analysis are presented in supplementary Table. Multivariate analyses demonstrated that the genetic and epidemiology factors were independent risk factors of leprosy (Table 3). Comparing with those in the control group, the individuals in the case group were more likely to be younger, male, minority, and had higher wGRS. Moreover, their index cases had higher baseline delay period of detection, and they were likely to be lepromatous leprosy patients.

The estimated C-index of every single risk factor's model showed poor predictive ability. Inclusion of all epidemiological factors did improve the predictive ability of the model. However, the combined model including all risk factors leads to a much greater C-index increase in the training cohort study (Table 4). The independently associated risk factors were used to form a leprosy risk estimation nomogram (Fig 2a). The C-index for leprosy prediction was 0·792 (95% CI 0·768 to 0·817), and was confirmed to be 0·780 through bootstrapping validation. The calibration plot for the risk of leprosy showed an optimal agreement between the prediction by nomogram and actual observation (Fig 2b). The Hosmer–Lemeshow test yielded a nonsignificant statistic (P = 0·819). The optimal cutoff value of the total nomogram scores was determined to be 81. HHCs were then divided into the low–risk group (score ≤ 81) and the high–risk group (score > 81) for further analysis. The nomogram classified 730 (66·4%) of 1100 contacts into the low-risk group and 370 (33·6%) contacts into the high–risk group. 406 (68·9%) of 589 leprosy cases were categorized into high–risk group while 487 (75·3%) of 647 healthy controls were categorized into low–risk group by the model. The odds ratio for leprosy between the high and low-risk group was 6·753 (95% CI 5·261-8·668, p < 0·001). The sensitivity and specificity of the nomogram were observed to be 68·9% (95% CI 65·0%-72·6%) and 75·3% (95% CI 71·8%-78·6%), respectively.

In the validation cohort, the follow-up time was 63 months. Among the 1100 leprosy HHCs, 12 subsequent leprosy cases were detected. The nomogram displayed a C-index of 0·773 (95%CI 0·658-0·888) for the estimation of leprosy risk. The Hosmer–Lemeshow test yielded a nonsignificant statistic (P = 0·511). The nomogram classified 730 (66·4%) of 1100 contacts into the low-risk group and 370 (33·6%) contacts into the high–risk group. Nine of 12 subsequent leprosy cases were categorized into high–risk group while 727 of 1088

Table 3
Multivariate Logistic Regression Analysis for Associations between Prediction Factors and Leprosy in the Training Cohort[a].

	OR (95% CI)	P value
Contact Age(years)		
11-20vs≤10	0.13 (0.06-0.25)	<0.001
21-30 vs≤10	0.09 (0.05-0.18)	<0.001
31-40 vs≤10	0.08 (0.04-0.14)	<0.001
41-50 vs≤10	0.06 (0.03-0.12)	<0.001
51-60 vs≤10	0.02 (0.01-0.05)	<0.001
61-70 vs≤10	0.02 (0.01-0.05)	<0.001
≥71 vs≤10	0.13 (0.02-0.73)	0.02
Ethnicity, Minority vs Han	1.36 (1.05-1.78)	0.02
Sex, Female vs Male	0.64 (0.49-0.83)	<0.001
Index cases' delay period of detection(months), >32.5 vs ≤32.5	2.54 (1.94-3.35)	<0.001
wGRS, >0.54 vs ≤0.54	2.07 (1.54-2.78)	<0.001
Ridley-Jopling Classification of Index cases		
BL vs BB	2.01 (1.20-3.45)	0.01
BT vs BB	0.97 (0.51-1.86)	0.93
I vs BB	2.08 (0.51-8.69)	0.30
LL vs BB	3.18 (1.90-5.47)	<0.001
TT vs BB	0.88 (0.47-1.67)	0.70

[a] Final model of multilevel logistic regression excluded individuals with missing data. OR, odds ratio; CI, confidence interval; TT, Tuberculoid Leprosy; BT, Borderline tuberculoid Leprosy; BB, Mid-borderline Leprosy; BL, Borderline lepromatous Leprosy; LL, Lepromatous Leprosy; I, Indeterminate Leprosy; wGRS, Weighted genetic risk score.

Table 4
Performance of risk models for leprosy.

Risk models	C-index (95%CI)
Contact age	0.698 (0.670, 0.726)
Ethnicity	0.532 (0.505, 0.559)
Sex	0.578 (0.551, 0.605)
Index cases' delay period of detection	0.601 (0.575, 0.627)
wGRS	0.566 (0.542, 0.591)
Ridley-Jopling Classification of Index cases	0.636 (0.607, 0.665)
Epidemiological risk factors [a]	0.783 (0.758, 0.808)
Epidemiological and genetic factors [b]	0.792 (0.768, 0.817)

[a] Epidemiological risk factors including contact age when the index cases were diagnosed, ethnicity, sex, index cases' delay period of detection, and Ridley-Jopling classification of index cases.
[b] Epidemiological and genetic factors including contact age when the index cases were diagnosed, ethnicity, sex, index cases' delay period of detection, Ridley-Jopling classification of index case, and wGRS. C-index, concordance index; CI, confidence interval; wGRS, Weighted genetic risk score.

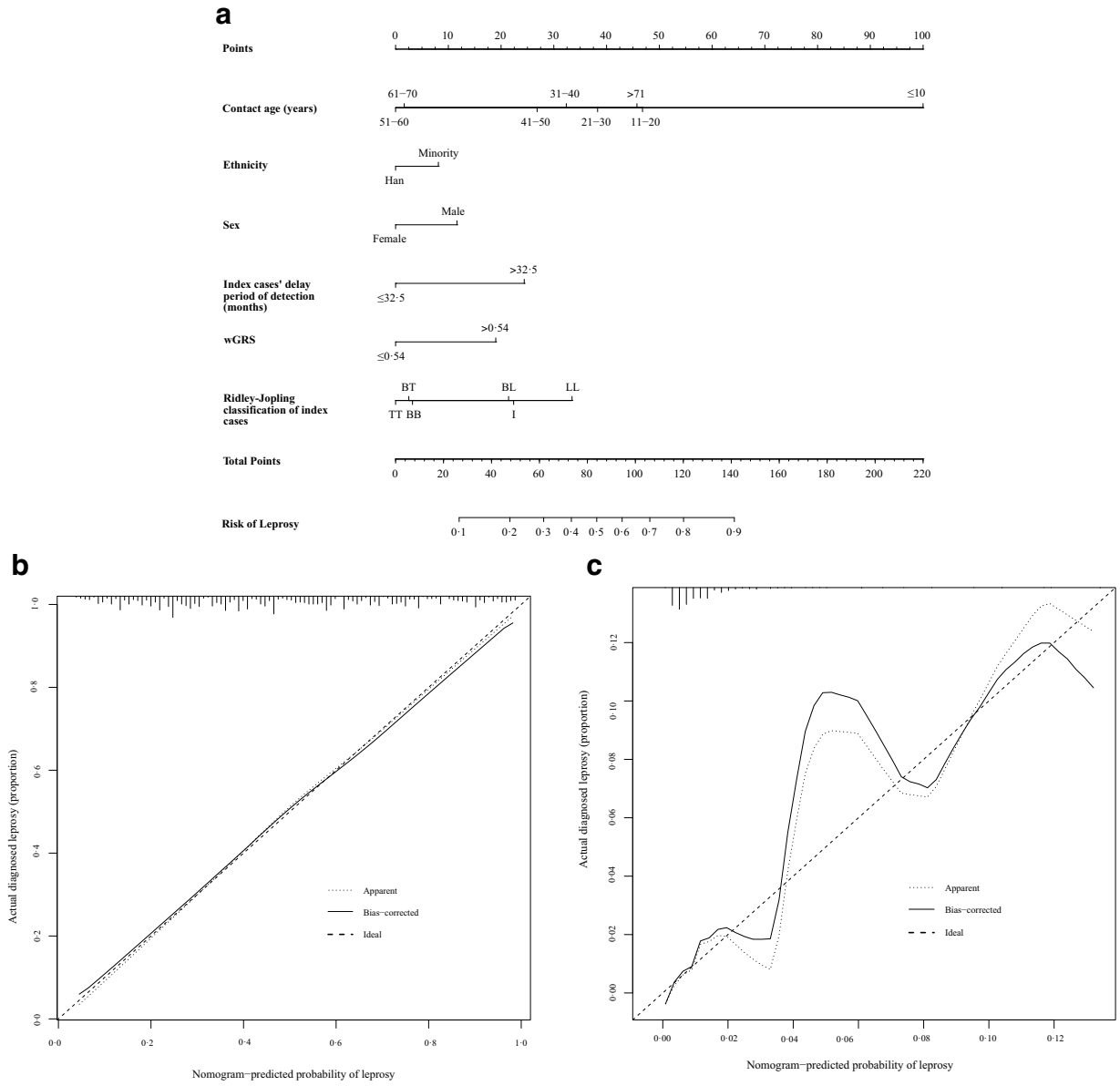

Fig. 2. Nomogram predicting leprosy probability in leprosy contacts and its predictive performance. a, to use the nomogram, an individual leprosy contact's value is located on each variable axis, and a line is drawn upward to determine the number of points received for each variable value. The sum of these numbers is located on the Total Points axis, and a line is drawn downward to the Risk of Leprosy axes to determine the leprosy probability. b, calibration curve of the nomogram prediction in the training cohort; c, calibration curve of the nomogram prediction in the validation cohort. The actual leprosy proportion is plotted on the y-axis; nomogram predicted probability is plotted on the x-axis. Model performance is shown by the Apparent line, relative to the 45-degree line, which represents perfect prediction. TT, Tuberculoid Leprosy; BT, Borderline tuberculoid Leprosy; BB, Mid-borderline Leprosy; BL, Borderline lepromatous Leprosy; LL, Lepromatous Leprosy; I, Indeterminate Leprosy.

healthy controls were categorized into low−risk group by the model. The odds ratio for leprosy between the high and low-risk group was 6·042 (95% CI 1·626-22·452, p = 0·006). The sensitivity and specificity of the nomogram were observed to be 75·0% (95% CI 42·8%-94·5%) and 66·8% (95% CI 63·9%-69·6%), respectively.

4. Discussion

In the current work based on a case−control study in the training cohort and an independent prospective cohort, we observed that risk factors including genetic factors (wGRS) and several epidemiological factors were significantly associated with leprosy.

Similar to our findings, some studies have suggested that lepromatous or multibacillary leprosy HHCs have a higher-risk than tuberculoid or paucibacillary leprosy HHCs [20], and higher odds of leprosy detection among contacts who are male [6]. Genetic factors can determine the immunologic response to *M. leprae*. A prospective cohort study demonstrated that genetic relationship is a relevant risk factor [5]. The wGRS was a proxy of genetic risk burden in our study, our results indicate that contacts with higher wGRS might be at higher-risk for developing leprosy. In this study, leprosy risk among HHCs was higher among the minority than Han residents, which might be explained by the different lifestyles, they prefer to live in their original area rather than work outside the home. We detected more cases of leprosy among contacts of index cases with longer delay period of detection. This finding might be explained by the exposure to the bacterium with relatively long period of time, and emphasizes the importance of leprosy early diagnosis and treatment. We studied the contact age when the index patient was diagnosed and found that it was an important risk predictor for the development of leprosy. The effect of age was bimodal, with an increased risk of age under 10 years and the risk increased again after age 60 years

which differed from previous studies [6,21]. It might be associated with the immune status of contact and the long latent period of *M. leprae*. Meanwhile, they may need more care from leprosy cases, resulting in more individuals' exposure to *M. leprae*. This reminds us to value the importance of HHCs' follow-up especially the old people and young children.

Genetic risk factors alone or in combination with clinical factors can be used for risk stratification and to guide strategies for treatment in various types of diseases [19,22]. A nomogram can provide an individualized, evidence-based, highly accurate risk estimation [17]. In the field of infectious diseases, a nomogram was constructed to predict severe coronavirus disease 2019, which could be helpful to better centralized management and early treatment of severe disease [23].

Several studies have reported some risk factors for the development of leprosy among HHCs [21]. However, a nomogram could present a quantitative and practical prediction tool for risk stratification. A previous study [24] that had no information regarding individuals' exposure to *M. leprae* attempted to use wGRS based on 25 variants to predict leprosy, the C-index in the training and validation cohort of this study were 0·773 and 0·707, respectively. Our nomogram was based on the information of leprosy HHCs and combined with wGRS and non-genetic factors has a significantly higher C-index in the training and validation cohort than it. (0·792/0·773 vs 0·743/0·707, respectively).

Furthermore, we evaluated the predictive accuracy of every single risk factor and improved the discriminatory ability of leprosy risk models by including genetic and epidemiological risk factors. We constructed a nomogram to predict leprosy risk and it performed well as supported by the C- index values of 0·792 and 0·773 in the training and validation cohorts, respectively. Meanwhile, the calibration curves demonstrated the agreements between prediction and actual observation in the training cohort. Owing to the limited subsequent leprosy cases, the calibration plot for the predicting outcome did not perfectly correspond to the actual outcome in the follow-up cohort (Fig 2c). However, the model showed good fit following Hosmer–Lemeshow test evaluation (P = 0·511). This nomogram had significantly high sensitivity and specificity to distinguish individuals with higher-risk of developing leprosy from HHCs based on a follow-up cohort using 81 as the cutoff value.

For the clinical use of the model, we summarized the sensitivity and the specificity in estimating the risk of leprosy using 81 as the cutoff value. The high-risk subgroup had a 6·042 times higher risk than the low-risk subgroup. This cutoff value may lead to a slight increase of false-positive rate according to the prospective study. However, in the setting of potential problems such as deformity caused by leprosy, a few high false-positive rates are acceptable. Considering the long incubation period of leprosy, the tracing time of this high-risk contact group needs to be extended. The predictive model can be used to identify contact individuals at a higher risk of developing leprosy. Meanwhile, this model may assist medical staff in the control of leprosy by better tracing higher-risk contact individuals such as providing more frequent visits and disease education.

The strengths of our study are that we validated the nomogram based on a prospective study to guarantee the robustness of the conclusion, and our study subjects were leprosy HHCs with more profound clinical significance. This study has some limitations. First, the analysis was based on data from high-endemic area, and the genetic variants were all identified in the Chinese Han population. Thus, the results should be validated from other endemic populations with larger prospective contacts cohorts and it might be necessary to study more polymorphisms associated with leprosy. Second, the retrospective design in the training cohort has a potential for bias. For example, in the training cohort, the time from index cases diagnosis to subsequent patients' onset or to the end of follow-up in control group is difficult to balance. Third, overcrowding, poor socioeconomic conditions [25] and antibodies to the *M. leprae* (phenolic glycolipid I antigen, leprosy IDRI diagnostic-1 and major membrane protein -Ⅱ) among leprosy contact subjects have been reported to be important for the development of leprosy [26,27] However, all of these risk factors were not included in the model as following reasons: leprosy cases and control individuals in the training cohort were matched according to their socioeconomic status and environmental conditions, meanwhile, serological markers titres showed changes during the leprosy progression [28], information regarding the antibody titres to specific *M. leprae* antigens when index cases were diagnosed in training cohort was missing. Thus, further prospective study should be done to investigate more non-genetic risk factors associated with leprosy. Finally, due to the low incidence of leprosy, only a small number of individuals in the validation cohort (<100) went on to develop the infection, we were unable to precisely estimate the model's performance and carry out the association analysis of risk factors [29].

5. Conclusion

Our data suggest that our nomogram can predict leprosy contacts at higher-risk of developing leprosy. The model is valuable for risk stratification management, which will be helpful for controlling leprosy transmission by formulating strategies for better targeting of PEP or tracing contacts.

Contributors

Conceptualization: H-SW; Data curation: S-YL, HL, LW, H-QJ, YS, W-YZ, J-SX, Y-QC, Y-MM, CP, Z-ZW, Z-WW; Formal analysis: S-YL, J-YS; Funding acquisition: H-SW; Investigation; M-WY, HL, LW and P-WS; Methodology: S-YL, J-YS and H-SW; Project administration: H-SW and M-WY; Resources: H-SW and M-WY; Software: S-YL and J-YS; Supervision: H-SW; Validation: H-SW; Visualization: S-YL and J-YS; Writing - original draft: S-YL; Writing - review & editing: H-SW and A-PW. All authors read and approved the final version of the manuscript.

Data Sharing Statement

All data, models, code generated or used during the study are available from the corresponding author upon reasonable request.

Declaration of Competing Interest

All authors declare that they have no conflicts of interest.

Acknowledgments

We thank all the staff who have been involved in the generation of data at each local institute. This work was supported by the Ministry of Health of China-research special funds for public health projects (grant number: 201502008), National Science and Technology Major Project (2018ZX10101-001), Chinese Academy of Medical Sciences Innovation Fund for Medical Science [2016-I2M-1-005, 2017-I2M-B&R-14], Jiangsu Provincial Science and Technology Project (BE2018619), The Nanjing Incubation Program for National Clinical Research Center (2019060001).

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2021.103408.

References

[1] Scollard DM, Adams LB, Gillis TP, et al. The continuing challenges of leprosy. Clin Microbiol Rev 2006;19(2):338–81.
[2] Global leprosy update, 2018: moving towards a leprosy- free world. 'Available at:' https://www.who.int/publications/i/item/who-wer9435-36. Accessed 16 February 2020.
[3] Chokkakula S, Shui T, Jiang H, et al. Genotyping of Mycobacterium leprae for understanding the distribution and transmission of leprosy in endemic provinces of China. Int J Infect Dis 2020;98:6–13.
[4] van Beers SM, Hatta M, Klatser PR. Patient contact is the major determinant in incident leprosy: implications for future control. Int J Lepr Other Mycobact Dis 1999;67(2):119–28.
[5] Moet FJ, Pahan D, Schuring RP, et al. Physical distance, genetic relationship, age, and leprosy classification are independent risk factors for leprosy in contacts of patients with leprosy. J Infect Dis 2006;193(3):346–53.
[6] Teixeira C, Pescarini JM, Alves F, et al. Incidence of and factors associated with leprosy among household contacts of patients with leprosy in Brazil. Jama Dermatol 2020;156(6):1–10.
[7] Pedrosa VL, Dias LC, Galban E, et al. Leprosy among schoolchildren in the Amazon region: A cross-sectional study of active search and possible source of infection by contact tracing. PLoS Negl Trop Dis 2018;12(2):e6261.
[8] Zhang F, Liu H, Chen S, et al. Genomewide association study of leprosy. N Engl J Med 2009;361(27):2609–18.
[9] Wang D, Fan Y, Malhi M, et al. Missense variants in HIF1A and LACC1 contribute to leprosy risk in Han Chinese. Am J Hum Genet 2018;102(5):794–805.
[10] Liu H, Wang Z, Li Y, et al. Genome-wide analysis of protein-coding variants in leprosy. J Invest Dermatol 2017;137(12):2544–51.
[11] Wang Z, Sun Y, Fu X, et al. A large-scale genome-wide association and meta-analysis identified four novel susceptibility loci for leprosy. Nat Commun 2016;7:13760.
[12] Liu H, Irwanto A, Fu X, et al. Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. Nat Genet 2015;47(3):267–71.
[13] Liu H, Irwanto A, Tian H, et al. Identification of IL18RAP/IL18R1 and IL12B as leprosy risk genes demonstrates shared pathogenesis between inflammation and infectious diseases. Am J Hum Genet 2012;91(5):935–41.
[14] Zhang F, Liu H, Chen S, et al. Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. Nat Genet 2011;43(12):1247–51.
[15] Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-group system. Int J Lepr Other Mycobact Dis 1966;34(3):255–73.
[16] Kidd AC, McGettrick M, Tsim S, et al. Survival prediction in mesothelioma using a scalable Lasso regression model: instructions for use and initial performance using clinical predictors. BMJ Open Respir Res 2018;5(1):e240.
[17] Lei Z, Li J, Wu D, et al. Nomogram for preoperative estimation of microvascular invasion risk in Hepatitis B virus-related hepatocellular carcinoma within the milan criteria. JAMA Surg 2016;151:356–63.
[18] Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited. Crit Care Med 2007;35(9):2052–6.
[19] Zhang JX, Song W, Chen ZH, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. Lancet Oncol 2013;14(13):1295–306.
[20] Richardus JH, Oskam L. Protecting people against leprosy: chemoprophylaxis and immunoprophylaxis. Clin Dermatol 2015;33(1):19–25.
[21] Feenstra SG, Nahar Q, Pahan D, et al. Social contact patterns and leprosy disease: a case-control study in Bangladesh. Epidemiol Infect 2013;141(3):573–81.
[22] Tang XR, Li YQ, Liang SB, et al. Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study. Lancet Oncol 2018;19(3):382–93.
[23] Gong J, Ou J, Qiu X, et al. A tool to early predict severe corona virus disease 2019 (COVID-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020;71(15):833–40.
[24] Wang N, Wang Z, Wang C, et al. Prediction of leprosy in the Chinese population based on a weighted genetic risk score. PLoS Negl Trop Dis 2018;12(9):e6789.
[25] Leano H, Araújo K, Bueno IC, Niitsuma E, Lana F. Socioeconomic factors related to leprosy: an integrative literature review. Rev Bras Enferm 2019;72(5):1405–15.
[26] Wang H, Liu W, Jin Y, et al. Detection of antibodies to both M. leprae PGL-I and MMP-II to recognize leprosy patients at an early stage of disease progression. Diagn Microbiol Infect Dis 2015;83(3):274–7.
[27] Amorim FM, Nobre ML, Ferreira LC, et al. Identifying Leprosy and Those at Risk of Developing Leprosy by Detection of Antibodies against LID-1 and LID-NDO. PLoS Negl Trop Dis 2016;10(9):e4934.
[28] Spencer JS, Duthie MS, Geluk A, et al. Identification of serological biomarkers of infection, disease progression and treatment efficacy for leprosy. Mem Inst Oswaldo Cruz 2012;107(Suppl 1):79–89.
[29] Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med 2016;35(2):214–26.