



Article

A Machine-Learning-Based Risk-Prediction Tool for HIV and Sexually Transmitted Infections Acquisition over the Next 12 Months

Xianglong Xu ^{1,2,3}, Zongyuan Ge ⁴, Eric P. F. Chow ^{1,2,5}, Zhen Yu ^{2,4}, David Lee ¹, Jinrong Wu ⁶, Jason J. Ong ^{1,2,3}, Christopher K. Fairley ^{1,2,3} and Lei Zhang ^{1,2,3,7,*}

- ¹ Melbourne Sexual Health Centre, Alfred Health, Melbourne, VIC 3053, Australia; xianglong.xu@monash.edu (X.X.); echow@mshc.org.au (E.P.F.C.); dlee@mshc.org.au (D.L.); doctorjasonong@gmail.com (J.J.O.); christopher.fairley@monash.edu (C.K.F.)
- ² Central Clinical School, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, VIC 3800, Australia; zhen.yu@monash.edu
- ³ China Australia Joint Research Center for Infectious Diseases, School of Public Health, Xi'an Jiaotong University Health Science Centre, Xi'an 710061, China
- ⁴ Monash e-Research Centre, Faculty of Engineering, Airdoc Research, Nvidia AI Technology Research Centre, Monash University, Melbourne, VIC 3800, Australia; zongyuan.ge@monash.edu
- ⁵ Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC 3053, Australia
- ⁶ Research Centre for Data Analytics and Cognition, La Trobe University, Bundoora, VIC 3086, Australia; melody.wu@latrobe.edu.au
- ⁷ Department of Epidemiology and Biostatistics, College of Public Health, Zhengzhou University, Zhengzhou 450001, China
- * Correspondence: lei.zhang1@monash.edu



Citation: Xu, X.; Ge, Z.; Chow, E.P.F.; Yu, Z.; Lee, D.; Wu, J.; Ong, J.J.; Fairley, C.K.; Zhang, L. A Machine-Learning-Based Risk-Prediction Tool for HIV and Sexually Transmitted Infections Acquisition over the Next 12 Months. *J. Clin. Med.* **2022**, *11*, 1818. <https://doi.org/10.3390/jcm11071818>

Academic Editor: Loredana Sarmati

Received: 2 March 2022

Accepted: 23 March 2022

Published: 25 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: More than one million people acquire sexually transmitted infections (STIs) every day globally. It is possible that predicting an individual's future risk of HIV/STIs could contribute to behaviour change or improve testing. We developed a series of machine learning models and a subsequent risk-prediction tool for predicting the risk of HIV/STIs over the next 12 months. Methods: Our data included individuals who were re-tested at the clinic for HIV (65,043 consultations), syphilis (56,889 consultations), gonorrhoea (60,598 consultations), and chlamydia (63,529 consultations) after initial consultations at the largest public sexual health centre in Melbourne from 2 March 2015 to 31 December 2019. We used the receiver operating characteristic (AUC) curve to evaluate the model's performance. The HIV/STI risk-prediction tool was delivered via a web application. Results: Our risk-prediction tool had an acceptable performance on the testing datasets for predicting HIV (AUC = 0.72), syphilis (AUC = 0.75), gonorrhoea (AUC = 0.73), and chlamydia (AUC = 0.67) acquisition. Conclusions: Using machine learning techniques, our risk-prediction tool has acceptable reliability in predicting HIV/STI acquisition over the next 12 months. This tool may be used on clinic websites or digital health platforms to form part of an intervention tool to increase testing or reduce future HIV/STI risk.

Keywords: HIV; sexually transmitted infections; machine learning; risk prediction; behavioural intervention

1. Introduction

HIV and sexually transmitted infections (STIs) are global public health concerns [1,2]. The World Health Organization (WHO) estimates that more than one million people acquire an STI every day. Given the rising rates of STIs, the WHO proposed the *Global health sector strategy on Sexually Transmitted Infections, 2016–2021*, to end STI epidemics as public health concerns by 2030, which included a 90% reduction in gonorrhoea incidence globally (2018 global baseline) and less than 50 cases of congenital syphilis per 100,000 live births in 80%

of countries [3]. In 2018, the 2030 agenda for sustainable development called for an end to the AIDS epidemic by 2030 [4]. One key strategy to reduce the incidence of HIV/STIs is to increase testing [5–7]. Barriers to testing include poor perception of HIV/STI risk, limited availability of testing, and cost [8]. Additionally, a delayed HIV diagnosis was also a common problem and caused numerous adverse health consequences [9,10]. Web-based apps for screening could effectively increase the uptake of health screening [11] and have usability and acceptability among users [12]. Predicting an individual's future risk of HIV/STIs could contribute to behaviour change or improve testing. To the best of our knowledge, no web-based prediction tool has yet been developed to predict an individual's risk of acquiring HIV/an STI over the next 12 months.

Machine learning algorithms have advantages for developing predictive models, such as not requiring statistical inferences or assumptions, being data driven, automatically learning from data that identifies complex nonlinear patterns, and exploiting complex interactions between risk factors [13]. Machine learning models have been used to predict the future risk of other conditions such as suicide [14,15], type 2 diabetes [16], Alzheimer's disease [17], and myocardial infarction [18]. Two studies using electronic health-record data from the USA reported that machine learning could accurately predict future HIV infection. A study in Massachusetts, USA, reported that models using routinely collected data from electronic health records (EHR) and machine learning could accurately predict the one-year risk of acquiring HIV [19]. Another study from Kaiser Permanente Northern California, USA, reported that by using machine learning, EHR-based HIV-risk models could accurately predict an incident HIV diagnosis within three years based on 81 predictors [20]. However, none of these models have been translated into a risk-prediction tool for predicting HIV over the next 12 months. Although a few studies have been conducted on future HIV prediction, no research has been published using machine learning methods to predict syphilis, gonorrhoea, and chlamydia acquisition over the next 12 months among males and females.

The purpose of this study was to use machine learning models, including bagging, boosting, and stacking algorithms [21], and routinely collected data in the clinical settings to predict HIV and three common STIs (syphilis, gonorrhoea, and chlamydia) acquisition over the next 12 months among males and females.

2. Materials and Methods

2.1. Study Data for 12-Month HIV/STI Risk-Prediction Tool Development

We used EHR data at the Melbourne Sexual Health Centre (MSHC) to develop and validate the machine learning model. The MSHC is the largest public sexual health centre in Melbourne, Australia. In the MSHC, individuals' demographic information, sexual practices, overseas sexual contact, and history of engaging in sex work are recorded at each visit using a computer-assisted self interview [22]. We used data from 2 March 2015 to 31 December 2019. We did not include data from 2020 because the COVID-19 epidemic could have changed the re-testing patterns and sexual practices of those attending the MSHC [23,24]. Transgender people and individuals aged below 18 years were excluded. The study was approved by the Alfred Hospital Ethics Committee, Australia (Project Number: 124/18). All methods were carried out following the relevant guidelines and regulations of the Alfred Hospital Ethics Committee.

A new diagnosis of HIV was based on serology and required a previous negative test. A diagnosis of syphilis was based on a clinician classifying the infection as early syphilis (primary, secondary, and early latent (<2 years)) using serology or a polymerase chain reaction (PCR). A diagnosis of gonorrhoea was based on a culture or a nucleic acid amplification test (NAAT) at one or more anatomical sites. A diagnosis of chlamydia was based on an NAAT at one or more anatomical sites. Our analysis included 65,043 consultations that had tested for HIV, 56,889 consultations for syphilis, 60,598 consultations for gonorrhoea, and 63,529 consultations for chlamydia. For the syphilis, gonorrhoea, and

chlamydia analysis, the detailed inclusions and exclusions are in Tables S1a,b. Details of the data-cleaning procedure are provided in the Supplementary File.

2.2. Predictors for 12-Month HIV/STI Risk Prediction

We extracted routinely collected data from the EHR, including self-reported questions at the first visit for each visit interval (described below). The feature selection was informed by the literature review, expert opinion, and previous work [25]. This baseline predictor data for modelling included gender, age (≥ 18 years old), country of birth, sexual practices (e.g., had sex with a sex partner in the last 12 months, number of sex partners in the last 12 months), condom use with sex partners in the last 12 months, pre-exposure prophylaxis (PrEP) use, presenting with STI symptoms, living with HIV (for STI prediction), and reported sexual contact with partners with an STI (gonorrhoea, chlamydia, or syphilis) (summarised in Tables 1 and S2).

Table 1. Characteristics (proportion or median value) of the included subjects stratified by HIV and STIs over the next 12 months.

Predictors	HIV		Syphilis		Gonorrhoea		Chlamydia	
	No n (%)	Yes n (%)	No n (%)	Yes n (%)	No n (%)	Yes n (%)	No n (%)	Yes n (%)
Gender								
Female	16,478 (25.4%)	1 (1.5%)	14,476 (25.8%)	12 (1.6%)	18,018 (31.9%)	298 (7.3%)	18,652 (31.6%)	687 (15.0%)
Male	48,499 (74.6%)	65 (98.5%)	41,663 (74.2%)	738 (98.4%)	38,521 (68.1%)	3761 (92.7%)	40,299 (68.4%)	3891 (85.0%)
Men who have sex with men								
No	5797 (12.0%)	1 (1.5%)	3854 (9.3%)	14 (1.9%)	5036 (13.1%)	55 (1.5%)	6713 (16.7%)	403 (10.4%)
Yes	42,702 (88.0%)	64 (98.5%)	37,809 (90.7%)	724 (98.1%)	33,485 (86.9%)	3706 (98.5%)	33,586 (83.3%)	3488 (89.6%)
Country of birth								
Australia	30,473 (46.9%)	29 (43.9%)	25,887 (46.1%)	355 (47.3%)	25,587 (45.3%)	2023 (49.8%)	27,081 (45.9%)	2112 (46.1%)
Overseas	31,978 (49.2%)	34 (51.5%)	28,099 (50.1%)	367 (48.9%)	28812 (51.0%)	1900 (46.8%)	29,684 (50.4%)	2310 (50.5%)
Missing	2526 (3.9%)	3 (4.5%)	2153 (3.8%)	28 (3.7%)	2140 (3.8%)	136 (3.4%)	2186 (3.7%)	156 (3.4%)
Age at consultation								
Median [IQR]	29.0 (25.0, 35.0)	30.5 (27.0, 43.0)	29.0 (25.0, 36.0)	30.0 (26.0, 37.0)	29.0 (25.0, 35.0)	29.0 (25.0, 34.0)	29.0 (25.0, 35.0)	28.0 (24.0, 34.0)
Current PrEP use								
No	62,195 (95.7%)	64 (97.0%)	53,496 (95.3%)	658 (87.7%)	53,998 (95.5%)	3656 (90.1%)	56,519 (95.9%)	4167 (91.0%)
Yes	2782 (4.3%)	2 (3.0%)	2643 (4.7%)	92 (12.3%)	2541 (4.5%)	403 (9.9%)	2432 (4.1%)	411 (9.0%)
Current sex worker								
No	57,383 (88.3%)	65 (98.5%)	49,068 (87.4%)	736 (98.1%)	49,458 (87.5%)	3902 (96.1%)	51,981 (88.2%)	4418 (96.5%)
Yes	7594 (11.7%)	1 (1.5%)	7071 (12.6%)	14 (1.9%)	7081 (12.5%)	157 (3.9%)	6970 (11.8%)	160 (3.5%)

Note: IQR: interquartile range.

2.3. Model Development and Training for Building a 12-Month HIV/STI Risk-Prediction Tool

We established a series of linear and nonlinear machine learning models that involved Regression Algorithms, including Multivariate Logistic Regression (MLR) and Elastic-Net Regression (ENR); Support Vector Machine Algorithms, including the Linear Support Vector Machines (without kernel extensions) (SVM (Linear)), SVM with a Polynomial Basis Kernel (Kernel SVM (Polynomial)), and SVM with a Radial Basis Function Kernel (Kernel SVM (RBF)); Bagging Ensemble Algorithms, including the Bagged Flexible Discriminant Analysis (Bagged FDA), Bagged Flexible Discriminant Analysis using Generalised Cross Validation (Bagged FDA using gCV Pruning), Bagged Multivariate Adaptive Regression Splines using Generalised Cross Validation (Bagged MARS using gCV Pruning), Random Forest (RF), and Conditional Inference Random Forest (CIRF); Boosting Ensemble Algorithms, including the Boosted Generalised Linear Model (Boosted GLM), Gradient Boosting Machines (GBM), and eXtreme Gradient Boosting (XGBoost). Based on our unpublished work, we also built a stacking model with 3 base models: ENR+GBM+RF. We also developed Naïve Bayes (NB), K-Nearest Neighbour (KNN), and multi-layer perceptron (MLP). MLR, ENR, GBM, RF, NB, MLP, and the stacking ensemble learning model (ENR+GBM+RF) were built using the *h2o* package. The bagged FDA, bagged FDA using gCV Pruning, and bagged MARS using gCV Pruning was built using the *earth* package. CIRF was built using the *party* package. The Boosted GLM was built using the *mboost* package. XGBoost was built using the *xgboost* package. KNN was built using the *kknn* package. The SVM(Linear) were built using the *e1071* package. The Kernel SVM (Polynomial) and Kernel SVM (RBF) were built using the *kernlab* package.

We used random-forest-based imputation to handle the missing data. The random-forest-based imputation was built using the *mice* package in R. Our machine learning models used a one-hot encoding scheme on the category variables. We used the nested cross validation (five outer folds, ten inner folds) method for the STI models to better estimate the generalisation error and solve the overfitting and selection bias caused by using a single dataset for the model selection and model training [26]. The external cross validation loop was repeated five times to solve the variance caused by the choice of the dataset to split. The prevalence of each of the four infections was below 10%, which means the data were imbalanced. Imbalanced data may cause either over-fitted or under-performed prediction results [27]. We used random under sampling in the training dataset to address the data imbalance to solve the class imbalance problem. Furthermore, an inner 10-fold CV loop was created for each model to select the tuning hyper-parameters for the maximised area under the ROC curve (AUC) on the training fold [28]. For the HIV models, we used an 80:20 random under-sampling split based on the outcome (HIV infection status) to create a training dataset and testing dataset for the analysis due to only 0.1% of consultations having a positive HIV result. All of the HIV models were trained using the training dataset with a ten-fold cross validation method and assessed the model performance on the testing dataset. Considering our datasets had data imbalances, the performance of the machine learning models was evaluated with the area under the receiver operating characteristic curve (AUC) and F1 score. Besides, we used the variable importance analysis of HIV, syphilis, gonorrhoea, and chlamydia to estimate the contribution of each of the predictors for the four infections.

The machine learning models and statistical analyses were conducted with R 3.6.1 and R Studio 1.2.5019. We used frequencies, percentages, the median, and the interquartile range (IQR) to present the descriptive analysis. We used Poisson regression to calculate incidence rates. We used MATLAB R 2019a (The Mathworks, Natick, MA, USA) to plot figures.

2.4. Twelve-Month HIV/STI Risk Estimate

We used the machine learning model output probability to calculate the HIV/STI risk over the next 12 months. Our machine learning models predicted the probability of HIV/an STI with a normalised distribution between the values 0 and 1. The model-predicted probability was calibrated to the actual prevalence level of the HIV/STI in the

following manner. First, we ranked the model-predicted probability for each individual in ascending order for the best-selected model. Second, we divided the model testing datasets into 200 probability subgroups. This generated 200 data points for each model-predicted probability and infection prevalence. The choice of 200 was arbitrary but ensured at least 100 individuals were included in each subgroup. Third, we fitted the data using a logistic function to provide a fitting curve for each model-predicted probability and infection prevalence. The calibration process was performed in MATLAB R2019a (details in the Supplementary Materials).

2.5. Establishment of the 12-Month HIV/STI Risk-Prediction Tool and Implementation of the Tool on a Web Server

According to the results of the variable importance analysis for all the variables, our final HIV/STI risk-prediction questionnaire was made up of the most important predictors. We used the AUC sensitivity and specificity to re-evaluate the model's performance. Additionally, we also used the AUC to compare the performance between the best machine learning model using all predictors and the best machine learning model using selected important predictors. Our machine-learning-based risk-prediction tool was developed as a web application using the Shiny R package. Details are in the Section 3 and Supplementary Materials.

3. Results

3.1. Characteristics of the 12-Month HIV/STI Risk-Prediction Tool Development Data

The proportion of consultations that tested positive over the next 12 months for each infection between 2 March 2015 and 31 December 2019 was: 0.10% (66/65,043) for HIV, 1.32% (750/56,889) for syphilis, 6.70% (4059/60,598) for gonorrhoea, and 7.21% (4578/63,529) for chlamydia. The median age of the individuals was 29.00 (IQR 24.00–43.00) for the four infection datasets (Table 1). Further details are provided in the Supplementary Materials (Table S2).

The incidence was 0.21 [95%CI: 0.17–0.27] per 100 person years (PY) for HIV, 3.42 [95%CI: 3.18–3.67] per 100 PY for syphilis, 17.56 [95%CI: 17.02–18.10] per 100 PY for gonorrhoea, and 18.50 [95%CI: 17.97–19.04] per 100 PY for chlamydia (Table S3). The Kaplan–Meier survival curves for each infection are shown in Figure S1.

3.2. Selecting the Best Machine Learning Model for 12-Month HIV/STI Risk-Prediction Tool

Of the 17 models, the receiver operating characteristic (ROC) curve that showed the best prediction models for HIV was the Boosted GLM (AUC = 0.73), for syphilis was the Boosted GLM (AUC = 0.76), for gonorrhoea was the ensemble Elastic-Net Regression (ENR)+ Gradient Boosting Machines (GBM)+ Random Forest (RF) (AUC = 0.73), and for chlamydia was the ensemble ENR+GBM+RF (AUC = 0.67). Details of the model-evaluation metrics are shown in the Tables S4–S19.

3.3. Selecting the Most Important Predictors for the 12-Month HIV/STI Risk-Prediction Tool

The results of the variable importance analysis showed the contribution of the predictors for HIV, syphilis, gonorrhoea, and chlamydia acquisition over the next 12 months. The variable importance varies between 0 and 1, with higher values indicating a stronger contribution to the prediction. We used the Boosted GLM variable importance analysis to identify the top predictive factors for HIV and the GBM variable importance analysis for syphilis, gonorrhoea, and chlamydia. Based on the variable importance analyses for HIV, syphilis, gonorrhoea, and chlamydia, the factors that contributed the most to predicting HIV/STIs over the next 12 months included age, gender, sex worker, men who had sex with men in the past 12 months (MSM), country of birth, contact with a chlamydia case, contact with a syphilis case, the number of casual male sexual partners in the past 12 months, condom use with male partners in the past 12 months, condom use with female partners in the past 12 months, drug use, PrEP use, sex overseas in the past 12 months, HIV infection,

past chlamydia, past gonorrhoea, past syphilis, past hepatitis B, past genital warts, and past other STIs (Figure 1).

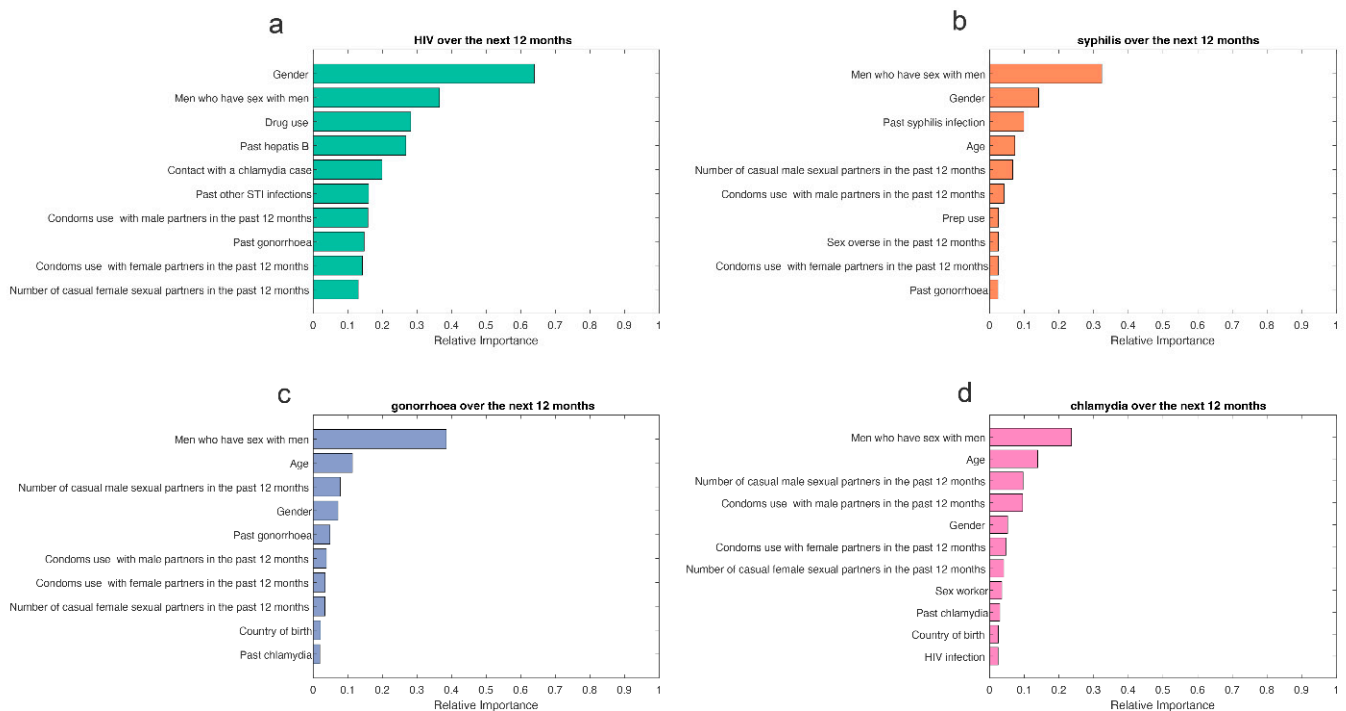


Figure 1. Variable importance analysis for predicting (a) HIV, (b) syphilis, (c) gonorrhoea, and (d) chlamydia over the next 12 months.

3.4. Establishment of the 12-Month HIV/STI Risk-Prediction Model

We built a risk-prediction model for HIV/STIs over the next 12 months using the most important predictors and the best model. Our risk-prediction model obtained an acceptable performance for predicting HIV (AUC = 0.72), syphilis (AUC = 0.75), gonorrhoea (AUC = 0.73), and chlamydia (AUC = 0.67), similar to its original model based on all the predictors (Figure 2, Tables S20 and S21). Details are shown in the Supplementary Materials.

3.5. Twelve-Month HIV/STI Risk Estimates and User Interface

To estimate the risk of twelve-month HIV/STIs, we fitted the data using a logistic function to provide a fitting curve for each model-predicted probability and infection prevalence (see Figures S2–S5). Details are shown in the Supplementary Materials. Our machine learning models were translated into a risk-prediction tool for predicting HIV/STIs. Our machine-learning-based risk-prediction tool was developed as a web application using the Shiny R package that creates the web-based tool named *MySTIRisk*. A prototype version of the tool is available at <https://ystirisk.shinyapps.io/mystirisk>, accessed on 1 March 2022. Figure 3 shows our proposed design for the user interface. The user interface has five modules: (1) the questionnaire survey module, (2) data-processing module, (3) HIV/STI risk prediction over the next 12 months, (4) testing recommendations, and (5) suggestions for risk reduction (Figure 3). Details are provided in the Supplementary File.

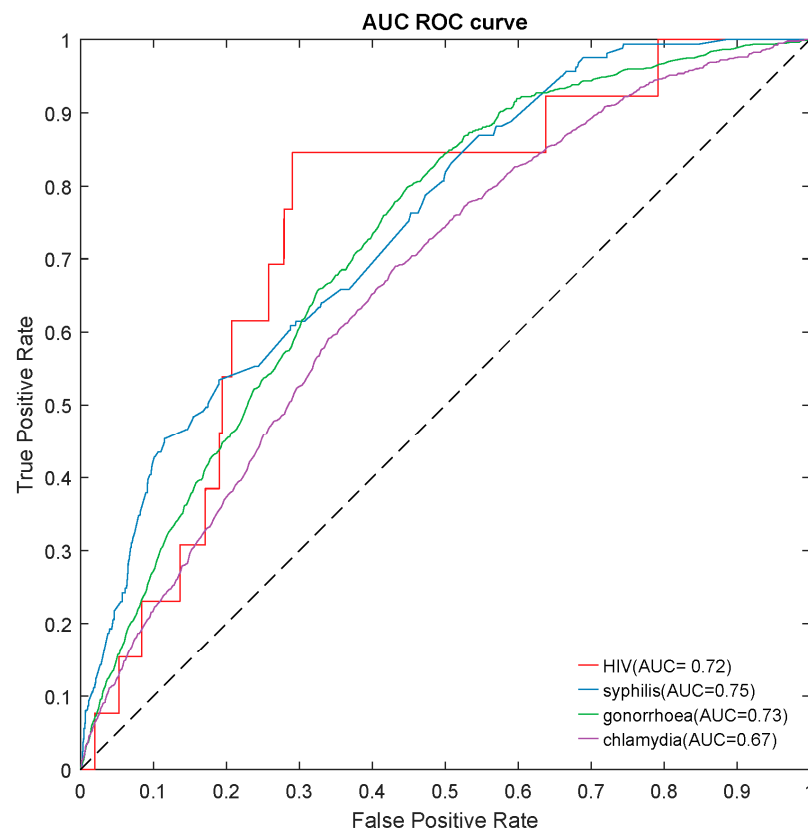


Figure 2. The area under the ROC curve (AUROC) of a risk-prediction tool for predicting HIV/STIs over the next 12 months on testing datasets. STI: syphilis, gonorrhoea, and chlamydia.

MySTIRisk: HIV/STI Risk over the next 12 months

Predicting HIV, syphilis, gonorrhoea, and chlamydia with machine learning algorithms over the next 12 months

Displayed below are the risks of chlamydia, gonorrhoea, syphilis and HIV for a person and reports the risks as you have over the next 12 months.

Please remember your risk of HIV/STI may change over time.

PLEASE NOTE these may not be the same as your risks and will depend on your personal circumstances

A prototype version of the tool is available at, <https://mystirisk.shinyapps.io/mystirisk>

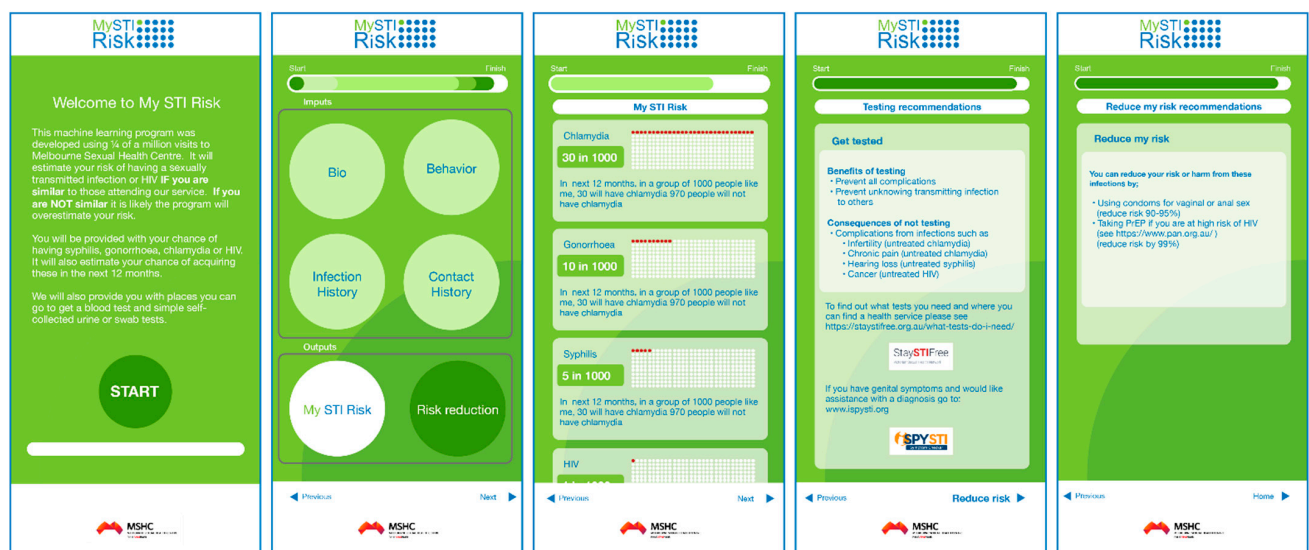


Figure 3. 12-month HIV/STI risk-prediction tool’s interface and output. STI: syphilis, gonorrhoea, and chlamydia.

4. Discussion

This is the first risk tool we are aware of that uses machine learning algorithms and routinely collects clinical data to predict the risk of acquiring HIV, syphilis, gonorrhoea, and chlamydia over the next 12 months. Our results showed that machine learning techniques could predict the risk of HIV and STIs over the next 12 months with acceptable reliability. Given that this tool uses routinely collected data and provides an immediate future risk, it has a number of potential applications. The potential applications include a web-based program for the public to assess their own future risk or to help clinical services triage high-risk individuals for further frequent screening or early public health interventions. Additional validation in other populations will be needed to evaluate the usefulness of this risk-prediction tool in other countries and regions. Future research should also focus on how best to communicate infection-risk information to the public and use it effectively to encourage them to increase testing or reduce risk and avoid over testing, anxiety, and false reassurance.

Risk prediction tools have been used as a part of interventions in other conditions, including COVID-19 [29,30], cardiovascular diseases [31,32], dementia [33], type 2 diabetes mellitus [34,35], cancer risk [12,36,37], autism [38], and falls [39]. These tools are generally well accepted by users in both public [31–36] and health professional domains, although they have mainly been used by health care professionals [30]. The interventions can result in an increased uptake of health information or services, such as screening [11]. For example, a large U.S. cohort used a web-based screening tool and substantially more participants sought information for their mental health [40]. Similarly, a screening app for mental health identified 159 patients from 733 users who were then advised to seek specialised care, of whom 55% started seeing a specialist [41]. Screening risk-assessment tools can also reduce unnecessary screening, as shown by a lung cancer tool that reduced the screening description among those ineligible for screening [37]. The use of apps to assess cardiovascular risk has been advocated as a method of identifying more at-risk individuals who can then access treatment within populations as an ‘add-on’ tool to enhance primary prevention [42]. These authors and others, such as the WHO, comment on the paucity of studies investigating the application of risk-assessment tools specifically directed to the public [43,44]. The studies described here highlight the complexity of risk-assessment tools for the public and suggest that improving an individual’s risk perception may lead to better healthcare-seeking behaviour. In a similar vein, previous authors have confirmed that an increase in the risk perception of an STI will likely improve subsequent healthcare use, such as testing or screening [45].

Based on these previous works on risk-assessment tools for other conditions, it is likely that our web-based HIV/STI risk-prediction tool may improve patient care, such as by improving access to sexual health care and increasing uptake and frequency of HIV/STI testing. For example, in California in the United States, a machine learning approach has been used to identify individuals at high risk of HIV and maybe a potential candidate for PrEP [20]. This information could be used to prompt clinicians to customise their intervention for the high-risk population in the clinical setting. However, recent reviews have indicated a relative lack of work relating to the use of AI in promoting HIV testing, and have attributed this lack to limited communication across the many different disciplines that are required for this type of research [46,47]. Individuals who use the tool may increase their HIV/STI risk perception and enable early screening or testing that is essential for HIV/STI prevention and control [48]. Our machine learning models identified some important predictors for HIV/STI acquisition over the next 12 months, consistent with previous studies. Previous research found various factors related to a high risk of incident HIV/STI, such as MSM [49], age (older for HIV and younger for STIs) [50], symptoms of or previous STIs [51], inconsistent condom use, PrEP use, and injecting drug use [52]. Providing our risk-prediction tool to these high-risk populations may improve the HIV/STI testing rate. However, we are also aware of the potential risk that an inappropriate interpretation of the risk score may lead some high-risk individuals to reduce their testing

or some low-risk individuals to possibly test inappropriately. It is also possible that the tool may lead to an increase in anxiety about HIV/STIs in some individuals. However, even for cancer risk assessment, this concern was relatively minor for the case of breast cancer risk prediction [12].

Our web-based HIV/STI risk-prediction tool may offer a useful method for potential sexual behavioural interventions to reduce future HIV/STI risk in addition to just promoting testing [53]. An example of this exists in cardiovascular risk where researchers used an individual's risk as part of an intervention for better lifestyle behaviours, including reducing smoking, more exercise, improving nutrition, and less stress [54]. The intervention group in this trial was shown to have more than two times a reduction in the Framingham scores for cardiovascular diseases than the control group over the next 12 months [54]. Therefore, in addition to potentially increasing testing, our HIV/STI risk-prediction tool could be incorporated into other preventive interventions, such as using PrEP. Such an addition would address one of the major challenges to increasing the PrEP uptake, which is identifying individuals who may benefit from HIV PrEP [55]. Nevertheless, we are aware of the possibility that providing risk scores and suggestions may not significantly change the targeted behaviour, as demonstrated by a randomised controlled trial study on cancer risk [56]. We recommend further controlled studies to examine if our HIV/STI risk-prediction tool would alter both short- and long-term behaviours.

This study has several limitations. First, the validity of the results depends on the accuracy of the self-reported information, which is subject to recall, non-response, and the social-desirability bias. Substantial work has been undertaken on our computer-assisted self-interviewing (CASI) system's validity and accuracy to ensure it performs well [57]. Second, the biggest challenge in developing our HIV risk-prediction model was the low incidence of HIV [19]. The HIV dataset had highly imbalanced data, with only 0.1% of the consultations having a positive HIV result. To address the problem of the limited HIV-positive samples in our machine learning training models, future machine learning models may employ more sophisticated machine learning techniques (e.g., transfer learning) [58], which may improve the accuracy of the models. In addition, the HIV data included a very small proportion of females, so our findings may not be generalisable to female users. Third, we used data from one clinic that services a population with a specific incidence of infection and demographic characteristics. This may not be representative of other population groups in the country or other global settings. Therefore, if users accessing the tool are not similar to those attending our clinical services, the risk estimate may be incorrect. However, one study comparing our clinic attendees and users accessing the MSHC website demonstrated similar characteristics and behaviour [59]. Further validation will be required if the prediction tool is used in other countries and regions. Fourth, the risks of HIV have changed rapidly over time and may continue to change. For example, the introduction of PrEP reduced HIV risk substantially, but condom use declined in the pre-exposure prophylaxis era [6]. Fifth, our models did not include data among individuals who tested positive on the day they conducted their questionnaire. This means that our estimated risk may be lower than it would have otherwise been. We did so to ensure that we measured the incidence of HIV/STI correctly. Sixth, the tool may be further improved by including more detailed behavioural information. For example, kissing and sequential sexual practices may contribute to gonorrhoea infection at more than one anatomical site [60]. Future HIV/STI predictive models may include these factors to improve the model's accuracy.

5. Conclusions

Our study demonstrates that EHR-based machine learning can predict HIV/STIs over the next 12 months. Based on the EHR in one of Australia's largest sexual health clinics, our web-based risk-assessment tool has an acceptable reliability in predicting the risk of HIV and three recurrent and asymptomatic STIs over the next 12 months. The risk-assessment tool can also be incorporated into a clinic to promote future HIV/STI testing or identify

individuals for HIV pre-exposure prophylaxis or early interventions for the reduction in future HIV/STI risk. Further validation studies in other countries can assess the usefulness of this risk-assessment tool, which helps reduce HIV/STI incidence and the cost of HIV/STI screening that requires expensive equipment and specialised expertise.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jcm11071818/s1>, Figure S1. The estimated survival curves for HIV and sexually transmitted infections using Kaplan-Meier. sexually transmitted infections: syphilis, gonorrhoea, and chlamydia. Kaplan-Meier survival curves stratified by risk groups over 365 days. Figure S2. The fitting curve of model-predicted probability and HIV prevalence over the next 12 months. Figure S3. The fitting curve of model-predicted probability and syphilis prevalence over the next 12 months. Figure S4. The fitting curve of model-predicted probability and gonorrhoea prevalence over the next 12 months. Figure S5. The fitting curve of model-predicted probability and chlamydia prevalence over the next 12 months. Table S1. (a) Inclusion and exclusion of HIV data. (b) Inclusion and exclusion of syphilis, gonorrhoea, and chlamydia data. Table S2. Characteristics (proportion or median value) of the included subjects stratified by HIV and STIs over the next 12 months. STIs: sexually transmitted infections. Table S3. Incidence of HIV and sexually transmitted infections per 100 person-years with 95% confidence intervals. sexually transmitted infections: syphilis, gonorrhoea, and chlamydia. Table S4. The area under ROC curve (AUC) of all predictors for predicting HIV over the next 12 months on testing data. ROC curve: receiver operating characteristic curve. Table S5. The area under ROC curve (AUC) of all predictors for predicting syphilis over the next 12 months on testing data. ROC curve: receiver operating characteristic curve. Table S6. The area under ROC curve (AUC) of all predictors for predicting gonorrhoea over the next 12 months on testing data. ROC curve: receiver operating characteristic curve. Table S7. The area under ROC curve (AUC) of all predictors for predicting chlamydia over the next 12 months on testing data. ROC curve: receiver operating characteristic curve. Table S8. Sensitivity of all predictors for predicting HIV over the next 12 months on testing data. Table S9. Sensitivity of all predictors for predicting syphilis over the next 12 months on testing data. Table S10. Sensitivity of all predictors for predicting gonorrhoea over the next 12 months on testing data. Table S11. Sensitivity of all predictors for predicting chlamydia over the next 12 months on testing data. Table S12. Specificity of all predictors for predicting HIV over the next 12 months on testing data. Table S13. Specificity of all predictors for predicting syphilis over the next 12 months on testing data. Table S14. Specificity of all predictors for predicting gonorrhoea over the next 12 months on testing data. Table S15. Specificity of all predictors for predicting chlamydia over the next 12 months on testing data. Table S16. F1 of all predictors for predicting HIV over the next 12 months on testing data. Table S17. F1 of all predictors for predicting syphilis over the next 12 months on testing data. Table S18. F1 of all predictors for predicting gonorrhoea over the next 12 months on testing data. Table S19. F1 of all predictors for predicting chlamydia over the next 12 months on testing data. Table S20. Performance metrics of the 12-month HIV/STI risk prediction tool (Best machine learning models using selected predictors). STI: syphilis, gonorrhoea, and chlamydia. Table S21. The performance comparison of the best machine learning models using all predictors and risk prediction tool. References [26–28,50,61] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, X.X., C.K.F. and L.Z.; methodology, X.X., Z.G., Z.Y. and L.Z.; software, X.X. and J.W.; validation, X.X., C.K.F. and L.Z.; formal analysis, X.X.; investigation, L.Z.; resources, C.K.F. and L.Z.; data curation, X.X., C.K.F. and L.Z.; writing—original draft preparation, X.X.; writing—review and editing, all authors; visualization, X.X. and J.W.; supervision, E.P.F.C., C.K.F. and L.Z.; project administration, L.Z.; funding acquisition, E.P.F.C., J.J.O., C.K.F. and L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: E.P.F.C. and J.J.O. are supported by an Australian National Health and Medical Research Council Emerging Leadership Investigator Grant (GNT1172873, GNT1193955, respectively). C.K.F. is supported by an Australian National Health and Medical Research Council Leadership Investigator Grant (GNT1172900). L.Z. is supported by the National Natural Science Foundation of China (Grant number: 81950410639); Outstanding Young Scholars Support Program (Grant number: 3111500001); Xi'an Jiaotong University Basic Research and Profession Grant (Grant number: xtr022019003, xzy032020032); Epidemiology modeling and risk assessment (Grant number: 20200344) and Xi'an Jiaotong University Young Scholar Support Grant (Grant number: YX6J004). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Institutional Review Board Statement: This study was approved by the Alfred Hospital Ethics Committee, Australia (Project Number: 124/18). All methods were carried out following the relevant guidelines and regulations of the Alfred Hospital Ethics Committee.

Informed Consent Statement: As this was a retrospective study involving minimal risk to the privacy of the study subjects, informed consent was waived by the Alfred Hospital Ethics Committee. All identifying details of the study subjects were removed before any computational analysis.

Data Availability Statement: The data is not publicly available due to privacy or ethical restrictions but will be made available on reasonable request from the corresponding author, with the permission of the Alfred Hospital Ethics Committee. Restrictions apply to the availability of the data used under the license for this study.

Acknowledgments: The authors would like to acknowledge Afrizal Afrizal from the Melbourne Sexual Health Centre for data extraction. The authors thank Glenda Fehler for her contribution to the data cleaning. The authors would also like to acknowledge Jon Emery from the University of Melbourne for an insightful discussion on risk-prediction tools (e.g., Figure 3). We thank Mark Chung at the MSHC for his assistance in preparing Figure 3.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ramchandani, M.S.; Golden, M.R. Confronting Rising STIs in the Era of PrEP and Treatment as Prevention. *Curr. HIV/AIDS Rep.* **2019**, *16*, 244–256. [[CrossRef](#)] [[PubMed](#)]
2. Institute of Medicine Committee on Prevention and Control of Sexually Transmitted Diseases. *The Hidden Epidemic: Confronting Sexually Transmitted Diseases*; Eng, T.R., Butler, W.T., Eds.; National Academies Press (US), National Academy of Sciences: Washington, DC, USA, 1997.
3. World Health Organization. Global health sector strategy on sexually transmitted infections 2016–2021: Toward ending STIs. In *Global Health Sector Strategy on Sexually Transmitted Infections 2016–2021: Toward Ending STIs*; WHO: Geneva, Switzerland, 2016.
4. UNAIDS. UNAIDS DATA 2018. Available online: https://www.unaids.org/sites/default/files/media_asset/unaids-data-2018_en.pdf (accessed on 1 March 2022).
5. Wei, C.; Herrick, A.; Raymond, H.F.; Anglemyer, A.; Gerbase, A.; Noar, S.M. Social marketing interventions to increase HIV/STI testing uptake among men who have sex with men and male-to-female transgender women. *Cochrane Database Syst. Rev.* **2011**, Cd009337. [[CrossRef](#)] [[PubMed](#)]
6. Chow, E.P.F.; Grulich, A.E.; Fairley, C.K. Epidemiology and prevention of sexually transmitted infections in men who have sex with men at risk of HIV. *Lancet HIV* **2019**, *6*, e396–e405. [[CrossRef](#)]
7. World Health Organization; Regional Office for South-East Asia. *Moving Ahead on Elimination of Sexually Transmitted Infections (STIs) in WHO South-East Asia Region—Progress and Challenges*; World Health Organization, Regional Office for South-East Asia: New Delhi, India, 2019.
8. Vermund, S.H.; Wilson, C.M. Barriers to HIV testing—where next? *Lancet* **2002**, *360*, 1186–1187. [[CrossRef](#)]
9. Lemoh, C.; Guy, R.; Yohannes, K.; Lewis, J.; Street, A.; Biggs, B.; Hellard, M. Delayed diagnosis of HIV infection in Victoria 1994 to 2006. *Sex. Health* **2009**, *6*, 117–122. [[CrossRef](#)] [[PubMed](#)]
10. Sobrino-Vegas, P.; Miguel, L.G.-S.; Caro-Murillo, A.M.; Miró, J.M.; Viciano, P.; Tural, C.; Saumoy, M.; Santos, I.; Sola, J.; Amo, J.d. Delayed diagnosis of HIV infection in a multicenter cohort: Prevalence, risk factors, response to HAART and impact on mortality. *Curr. HIV Res.* **2009**, *7*, 224–230. [[CrossRef](#)]
11. Ooi, C.Y.; Ng, C.J.; Sales, A.E.; Lim, H.M. Implementation Strategies for Web-Based Apps for Screening: Scoping Review. *J. Med. Internet Res.* **2020**, *22*, e15591. [[CrossRef](#)]
12. Lo, L.L.; Collins, I.M.; Bressel, M.; Butow, P.; Emery, J.; Keogh, L.; Weideman, P.; Steel, E.; Hopper, J.L.; Trainer, A.H.; et al. The iPrevent Online Breast Cancer Risk Assessment and Risk Management Tool: Usability and Acceptability Testing. *JMIR Form. Res.* **2018**, *2*, e24. [[CrossRef](#)]

13. Patel, B.; Sengupta, P. Machine learning for predicting cardiac events: What does the future hold? *Expert Rev. Cardiovasc. Ther.* **2020**, *18*, 77–84. [[CrossRef](#)]
14. Roy, A.; Nikolitch, K.; McGinn, R.; Jinah, S.; Klement, W.; Kaminsky, Z.A. A machine learning approach predicts future risk to suicidal ideation from social media data. *NPJ Digit. Med.* **2020**, *3*, 78. [[CrossRef](#)]
15. Whiting, D.; Fazel, S. How accurate are suicide risk prediction models? Asking the right questions for clinical practice. *Evid.-Based Ment. Health* **2019**, *22*, 125–128. [[CrossRef](#)] [[PubMed](#)]
16. Farran, B.; AlWotayan, R.; Alkandari, H.; Al-Abdulrazzaq, D.; Channanath, A.; Thanaraj, T.A. Use of Non-invasive Parameters and Machine-Learning Algorithms for Predicting Future Risk of Type 2 Diabetes: A Retrospective Cohort Study of Health Data from Kuwait. *Front. Endocrinol.* **2019**, *10*, 624. [[CrossRef](#)] [[PubMed](#)]
17. Park, J.H.; Cho, H.E.; Kim, J.H.; Wall, M.M.; Stern, Y.; Lim, H.; Yoo, S.; Kim, H.S.; Cha, J. Machine learning prediction of incidence of Alzheimer’s disease using large-scale administrative health data. *NPJ Digit. Med.* **2020**, *3*, 46. [[CrossRef](#)] [[PubMed](#)]
18. Kwiecinski, J.; Tzolos, E.; Meah, M.; Cadet, S.; Adamson, P.D.; Grodecki, K.; Joshi, N.V.; Moss, A.J.; Williams, M.C.; van Beek, E.J.; et al. Machine-learning with (18)F-sodium fluoride PET and quantitative plaque analysis on CT angiography for the future risk of myocardial infarction. *J. Nucl. Med.* **2021**, *63*, 158–165. [[CrossRef](#)]
19. Gruber, S.; Krakower, D.; Menchaca, J.T.; Hsu, K.; Hawrusik, R.; Maro, J.C.; Cocoros, N.M.; Kruskal, B.A.; Wilson, I.B.; Mayer, K.H.; et al. Using electronic health records to identify candidates for human immunodeficiency virus pre-exposure prophylaxis: An application of super learning to risk prediction when the outcome is rare. *Stat. Med.* **2020**, *39*, 3059–3073. [[CrossRef](#)]
20. Marcus, J.L.; Hurley, L.B.; Krakower, D.S.; Alexeeff, S.; Silverberg, M.J.; Volk, J.E. Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: A modelling study. *Lancet HIV* **2019**, *6*, e688–e695. [[CrossRef](#)]
21. Bzdok, D.; Altman, N.; Krzywinski, M. Statistics versus machine learning. *Nat. Methods* **2018**, *15*, 233–234. [[CrossRef](#)]
22. Misson, J.; Chow, E.P.F.; Chen, M.Y.; Read, T.R.H.; Bradshaw, C.S.; Fairley, C.K. Trends in gonorrhoea infection and overseas sexual contacts among females attending a sexual health centre in Melbourne, Australia, 2008–2015. *Commun. Dis. Intell.* **2018**, *42*, 1–10.
23. Chow, E.P.F.; Hocking, J.S.; Ong, J.J.; Phillips, T.R.; Fairley, C.K. Sexually Transmitted Infection Diagnoses and Access to a Sexual Health Service Before and After the National Lockdown for COVID-19 in Melbourne, Australia. *Open Forum Infect. Dis.* **2021**, *8*, ofaa536. [[CrossRef](#)]
24. Chow, E.P.F.; Ong, J.J.; Donovan, B.; Foster, R.; Phillips, T.R.; McNulty, A.; Fairley, C.K. Comparing HIV Post-Exposure Prophylaxis, Testing, and New Diagnoses in Two Australian Cities with Different Lockdown Measures during the COVID-19 Pandemic. *Int. J. Environ. Res. Public Health* **2021**, *18*, 10814. [[CrossRef](#)]
25. Bao, Y.; Medland, N.A.; Fairley, C.K.; Wu, J.; Shang, X.; Chow, E.P.F.; Xu, X.; Ge, Z.; Zhuang, X.; Zhang, L. Predicting the diagnosis of HIV and sexually transmitted infections among men who have sex with men using machine learning approaches. *J. Infect.* **2021**, *82*, 48–59. [[CrossRef](#)] [[PubMed](#)]
26. Shehzad, A.; Rockwood, K.; Stanley, J.; Dunn, T.; Howlett, S.E. Use of Patient-Reported Symptoms from an Online Symptom Tracking Tool for Dementia Severity Staging: Development and Validation of a Machine Learning Approach. *J. Med. Internet Res.* **2020**, *22*, e20840. [[CrossRef](#)] [[PubMed](#)]
27. Menardi, G.; Torelli, N. Training and assessing classification rules with imbalanced data. *Data Min. Knowl. Discov.* **2014**, *28*, 92–122. [[CrossRef](#)]
28. Liao, X.; Kerr, D.; Morales, J.; Duncan, I. Application of Machine Learning to Identify Clustering of Cardiometabolic Risk Factors in U.S. Adults. *Diabetes Technol. Ther.* **2019**, *21*, 245–253. [[CrossRef](#)]
29. Clift, A.K.; Coupland, C.A.C.; Keogh, R.H.; Diaz-Ordaz, K.; Williamson, E.; Harrison, E.M.; Hayward, A.; Hemingway, H.; Horby, P.; Mehta, N.; et al. Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: National derivation and validation cohort study. *BMJ* **2020**, *371*, m3731. [[CrossRef](#)]
30. Liang, W.; Liang, H.; Ou, L.; Chen, B.; Chen, A.; Li, C.; Li, Y.; Guan, W.; Sang, L.; Lu, J.; et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. *JAMA Intern. Med.* **2020**, *180*, 1081–1089. [[CrossRef](#)]
31. Manuel, D.G.; Tuna, M.; Bennett, C.; Hennessy, D.; Rosella, L.; Sanmartin, C.; Tu, J.V.; Perez, R.; Fisher, S.; Taljaard, M. Development and validation of a cardiovascular disease risk-prediction model using population health surveys: The Cardiovascular Disease Population Risk Tool (CVDPoRT). *CMAJ* **2018**, *190*, E871–E882. [[CrossRef](#)]
32. Rossello, X.; Dorresteijn, J.A.; Janssen, A.; Lambrinou, E.; Scherrenberg, M.; Bonnefoy-Cudraz, E.; Cobain, M.; Piepoli, M.F.; Visseren, F.L.; Dendale, P. Risk prediction tools in cardiovascular disease prevention: A report from the ESC Prevention of CVD Programme led by the European Association of Preventive Cardiology (EAPC) in collaboration with the Acute Cardiovascular Care Association (ACCA) and the Association of Cardiovascular Nursing and Allied Professions (ACNAP). *Eur. Heart J. Acute Cardiovasc. Care* **2020**, *9*, 522–532. [[CrossRef](#)]
33. Fisher, S.; Manuel, D.G.; Hsu, A.T.; Bennett, C.; Tuna, M.; Eddeen, A.B.; Sequeira, Y.; Jessri, M.; Taljaard, M.; Anderson, G.M.; et al. Development and validation of a predictive algorithm for risk of dementia in the community setting. *J. Epidemiol. Community Health* **2021**, *75*, 843–853. [[CrossRef](#)]
34. Lai, H.; Huang, H.; Keshavjee, K.; Guergachi, A.; Gao, X. Predictive models for diabetes mellitus using machine learning techniques. *BMC Endocr. Disord.* **2019**, *19*, 101. [[CrossRef](#)]

35. Lindström, J.; Tuomilehto, J. The diabetes risk score: A practical tool to predict type 2 diabetes risk. *Diabetes Care* **2003**, *26*, 725–731. [[CrossRef](#)]
36. Collins, I.M.; Bickerstaffe, A.; Ranaweera, T.; Maddumarachchi, S.; Keogh, L.; Emery, J.; Mann, G.B.; Butow, P.; Weideman, P.; Steel, E.; et al. iPrevent[®]: A tailored, web-based, decision support tool for breast cancer risk assessment and management. *Breast Cancer Res. Treat.* **2016**, *156*, 171–182. [[CrossRef](#)] [[PubMed](#)]
37. Lau, Y.K.; Caverly, T.J.; Cao, P.; Cherng, S.T.; West, M.; Gaber, C.; Arenberg, D.; Meza, R. Evaluation of a personalized, web-based decision aid for lung cancer screening. *Am. J. Prev. Med.* **2015**, *49*, e125–e129. [[CrossRef](#)] [[PubMed](#)]
38. Brooks, B.A.; Haynes, K.; Smith, J.; McFadden, T.; Robins, D.L. Implementation of web-based autism screening in an urban clinic. *Clin. Pediatr.* **2016**, *55*, 927–934. [[CrossRef](#)] [[PubMed](#)]
39. Poe, S.S.; Dawson, P.B.; Cvach, M.; Burnett, M.; Kumble, S.; Lewis, M.; Thompson, C.B.; Hill, E.E. The Johns Hopkins Fall Risk Assessment Tool: A Study of Reliability and Validity. *J. Nurs. Care Qual.* **2018**, *33*, 10–19. [[CrossRef](#)]
40. Jacobson, N.C.; Yom-Tov, E.; Lekkas, D.; Heinz, M.; Liu, L.; Barr, P.J. Impact of online mental health screening tools on help-seeking, care receipt, and suicidal ideation and suicidal intent: Evidence from internet search behavior in a large U.S. cohort. *J. Psychiatr. Res.* **2022**, *145*, 276–283. [[CrossRef](#)]
41. Diez-Canseco, F.; Toyama, M.; Ipince, A.; Perez-Leon, S.; Caverro, V.; Araya, R.; Miranda, J.J. Integration of a Technology-Based Mental Health Screening Program into Routine Practices of Primary Health Care Services in Peru (The Allillanchu Project): Development and Implementation. *J. Med. Internet Res.* **2018**, *20*, e100. [[CrossRef](#)]
42. Feigin, V.L.; Norrving, B.; Mensah, G.A. Primary prevention of cardiovascular disease through population-wide motivational strategies: Insights from using smartphones in stroke prevention. *BMJ Glob. Health* **2016**, *2*, e000306. [[CrossRef](#)]
43. Kay, M.; Santos, J.; Takane, M. mHealth: New horizons for health through mobile technologies. *World Health Organ.* **2011**, *64*, 66–71.
44. Turakhia, M.P.; Desai, S.A.; Harrington, R.A. The outlook of digital health for cardiovascular medicine: Challenges but also extraordinary opportunities. *JAMA Cardiol.* **2016**, *1*, 743–744. [[CrossRef](#)]
45. Clifton, S.; Mercer, C.H.; Sonnenberg, P.; Tanton, C.; Field, N.; Gravningen, K.; Hughes, G.; Mapp, F.; Johnson, A.M. STI risk perception in the British population and how it relates to sexual behaviour and STI healthcare use: Findings from a cross-sectional survey (Natsal-3). *EclinicalMedicine* **2018**, *2*, 29–36. [[CrossRef](#)] [[PubMed](#)]
46. Xiang, Y.; Du, J.; Fujimoto, K.; Li, F.; Schneider, J.; Tao, C. Application of artificial intelligence and machine learning for HIV prevention interventions. *Lancet HIV* **2022**, *9*, e54–e62. [[CrossRef](#)]
47. Marcus, J.L.; Sewell, W.C.; Balzer, L.B.; Krakower, D.S. Artificial Intelligence and Machine Learning for HIV Prevention: Emerging Approaches to Ending the Epidemic. *Curr. HIV/AIDS Rep.* **2020**, *17*, 171–179. [[CrossRef](#)] [[PubMed](#)]
48. World Health Organization. *Sexually Transmitted Infections (STIs): The Importance of a Renewed Commitment to STI Prevention and Control in Achieving Global Sexual and Reproductive Health*; World Health Organization: Geneva, Switzerland, 2013.
49. Garofalo, R.; Hotton, A.L.; Kuhns, L.M.; Gratz, B.; Mustanski, B. Incidence of HIV Infection and Sexually Transmitted Infections and Related Risk Factors Among Very Young Men Who Have Sex with Men. *J. Acquir. Immune Defic. Syndr.* **2016**, *72*, 79–86. [[CrossRef](#)] [[PubMed](#)]
50. Selvey, L.A.; Slimings, C.; Adams, E.; Manuel, J. Incidence and predictors of HIV, chlamydia and gonorrhoea among men who have sex with men attending a peer-based clinic. *Sex. Health* **2018**, *15*, 451–459. [[CrossRef](#)] [[PubMed](#)]
51. Dukers-Muijers, N.; van Rooijen, M.S.; Hogewoning, A.; van Liere, G.; Steenbakkens, M.; Hoebe, C. Incidence of repeat testing and diagnoses of Chlamydia trachomatis and Neisseria gonorrhoea in swingers, homosexual and heterosexual men and women at two large Dutch STI clinics, 2006–2013. *Sex. Transm. Infect.* **2017**, *93*, 383–389. [[CrossRef](#)] [[PubMed](#)]
52. Cheung, K.T.; Fairley, C.K.; Read, T.R.; Denham, I.; Fehler, G.; Bradshaw, C.S.; Chen, M.Y.; Chow, E.P. HIV Incidence and Predictors of Incident HIV among Men Who Have Sex with Men Attending a Sexual Health Clinic in Melbourne, Australia. *PLoS ONE* **2016**, *11*, e0156160. [[CrossRef](#)]
53. Lustria, M.L.A.; Noar, S.M.; Cortese, J.; Van Stee, S.K.; Glueckauf, R.L.; Lee, J. A meta-analysis of web-delivered tailored health behavior change interventions. *J. Health Commun.* **2013**, *18*, 1039–1069. [[CrossRef](#)]
54. Wister, A.; Loewen, N.; Kennedy-Symonds, H.; McGowan, B.; McCoy, B.; Singer, J. One-year follow-up of a therapeutic lifestyle intervention targeting cardiovascular disease risk. *CMAJ* **2007**, *177*, 859–865. [[CrossRef](#)]
55. Underhill, K.; Operario, D.; Skeer, M.; Mimiaga, M.; Mayer, K. Packaging PrEP to prevent HIV: An integrated framework to plan for pre-exposure prophylaxis implementation in clinical practice. *J. Acquir. Immune Defic. Syndr.* **2010**, *55*, 8–13. [[CrossRef](#)]
56. Yuwaki, K.; Kuchiba, A.; Otsuki, A.; Odawara, M.; Okuhara, T.; Ishikawa, H.; Inoue, M.; Tsugane, S.; Shimazu, T. Effectiveness of a Cancer Risk Prediction Tool on Lifestyle Habits: A Randomized Controlled Trial. *Cancer Epidemiol. Biomark. Prev.* **2021**, *30*, 1063–1071. [[CrossRef](#)] [[PubMed](#)]
57. Fairley, C.K.; Sze, J.K.; Vodstrcil, L.A.; Chen, M.Y. Computer-assisted self interviewing in sexual health clinics. *Sex. Transm. Dis.* **2010**, *37*, 665–668. [[CrossRef](#)] [[PubMed](#)]
58. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. *Proc. IEEE* **2020**, *109*, 43–76. [[CrossRef](#)]
59. Lee, D.M.; Fairley, C.K.; Sze, J.K.; Kuo, T.; Cummings, R.; Bilardi, J.; Chen, M.Y. Access to sexual health advice using an automated, internet-based risk assessment service. *Sex. Health* **2009**, *6*, 63–66. [[CrossRef](#)] [[PubMed](#)]

60. Xu, X.; Chow, E.P.F.; Ong, J.J.; Hoebe, C.; Williamson, D.; Shen, M.; Kong, F.Y.S.; Hocking, J.S.; Fairley, C.K.; Zhang, L. Modelling the contribution that different sexual practices involving the oropharynx and saliva have on *Neisseria gonorrhoeae* infections at multiple anatomical sites in men who have sex with men. *Sex. Transm. Infect.* **2021**, *97*, 183–189. [[CrossRef](#)]
61. Vandormael, A.; Dobra, A.; Bärnighausen, T.; de Oliveira, T.; Tanser, F. Incidence rate estimation, periodic testing and the limitations of the mid-point imputation approach. *Int. J. Epidemiol.* **2018**, *47*, 236–245. [[CrossRef](#)]