

RESEARCH ARTICLE

Personalizing the empiric treatment of gonorrhea using machine learning models

Rachel E. Murray-Watson^{1,2*}, Yonatan H. Grad³, Sancta B. St. Cyr⁴, Reza Yaesoubi^{1,2}

1 Department of Health Policy and Management, Yale School of Public Health, New Haven, Connecticut, United States of America, **2** Public Health Modeling Unit, Yale School of Public Health, New Haven, Connecticut, United States of America, **3** Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, United States of America, **4** Division of STD Prevention, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, Georgia, United States of America

* rachel.murray-watson@yale.edu

Abstract

Despite the emergence of antimicrobial-resistant (AMR) strains of *Neisseria gonorrhoeae*, the treatment of gonorrhea remains empiric and according to standardized guidelines, which are informed by the national prevalence of resistant strains. Yet, the prevalence of AMR varies substantially across geographic and demographic groups. We investigated whether data from the national surveillance system of AMR gonorrhea in the US could be used to personalize the empiric treatment of gonorrhea. We used data from the Gonococcal Isolate Surveillance Project collected between 2000–2010 to train and validate machine learning models to identify resistance to ciprofloxacin (CIP), one of the recommended first-line antibiotics until 2007. We used these models to personalize empiric treatments based on sexual behavior and geographic location and compared their performance with standardized guidelines, which recommended treatment with CIP, ceftriaxone (CRO), or cefixime (CFX) between 2005–2006, and either CRO or CFX between 2007–2010. Compared with standardized guidelines, the personalized treatments could have replaced 33% of CRO and CFX use with CIP while ensuring that 98% of patients were prescribed effective treatment during 2005–2010. The models maintained their performance over time and across geographic regions. Predictive models trained on data from national surveillance systems of AMR gonorrhea could be used to personalize the empiric treatment of gonorrhea based on patients' basic characteristics at the point of care. This approach could reduce the unnecessary use of newer antibiotics while maintaining the effectiveness of first-line therapy.

OPEN ACCESS

Citation: Murray-Watson RE, Grad YH, St. Cyr SB, Yaesoubi R (2024) Personalizing the empiric treatment of gonorrhea using machine learning models. PLOS Digit Health 3(8): e0000549. <https://doi.org/10.1371/journal.pdig.0000549>

Editor: Dhiya Al-Jumeily O. B. E., Liverpool John Moores University - City Campus: Liverpool John Moores University, UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

Received: January 24, 2024

Accepted: June 11, 2024

Published: August 14, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pdig.0000549>

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: The data underlying the results presented in the study are available

Author summary

Treating gonorrhea is complicated by the spread of drug-resistant strains, yet current approaches rely on empiric guidelines. Our study explored using national data on resistance in gonorrhea to personalize treatment decisions. From 2000 to 2010, we analyzed information from the Gonococcal Isolate Surveillance Project to develop predictive models. These models considered factors like where patients live and their behavior to suggest

from the Gonococcal Isolate Surveillance Project (<https://www.cdc.gov/std/gisp/default.htm>).

Funding: This work was supported by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health (R01AI153351 to RY). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

treatments tailored to their likely resistance profile. Compared to standard guidelines during that period, our personalized approach could have reduced unnecessary use of certain antibiotics by 33%, while still effectively treating 98% of patients. These findings highlight the potential of using data-driven models to improve how we treat gonorrhea, ensuring effective care while safeguarding newer antibiotics for the future.

Introduction

Antimicrobial-resistant (AMR) gonorrhea is a pressing issue. *Neisseria gonorrhoeae**, the causative agent of gonorrhea, has proven adept at evolving or acquiring resistance mechanisms to all antibiotics that have been used to treat it [1]. In 2019 alone, there were an estimated 942,000 cases of AMR gonococcal infections [2], an increase of 71.3% on the 2017 estimate. This makes it one of the most urgent antibiotic-resistant threats in the United States (US) [3].

Though the susceptibility of gonococcal strains to different antibiotics can be determined using bacterial cultures, they have mostly been replaced by nucleic acid amplification tests (NAATs) in the diagnosis of gonorrhea [4]. Without antimicrobial-specific diagnostic information, gonorrhea is treated empirically according to standardized treatment guidelines, which are determined at the national level and based on estimates for the prevalence of AMR gonorrhea. In the US, for example, these estimates are provided by the Gonococcal Isolate Surveillance Project (GISP), a national surveillance system that collects data from an average of 28 sentinel sites in the US [5].

To ensure the effectiveness of empiric first-line therapy, guidelines recommend antibiotics with a low national prevalence of resistance (usually <5%) for all patients with gonorrhea [6–8]. However, reports of substantial variation in the AMR prevalence and trend among different geographic regions or population subgroups [9–13]. Given the empiric nature of treatment, these variations led to large disparities in the number of patients receiving effective treatment. Such concerns resulted in revised guidelines that accounted for these variations. For example, in 2000, estimates from GISP indicated that there was an increase in the prevalence of ciprofloxacin-resistant (CIP-R) gonococcal infections in Hawai'i. In response, CDC revised the guidelines for the state of Hawai'i, no longer recommending CIP as the first-line treatment [6]. Two years later, in 2002, CIP was also removed from the treatment recommendations for gonorrhea in California. By 2004, the advice against the empirical use of CIP was extended to MSM, and finally, in the year 2007, the CDC recommended against using CIP for all patients with gonorrhea.

These refinements highlight the need for a principled approach to personalize empiric treatment based on patient characteristics that are associated with the risk of AMR gonorrhea (such as sexual behavior and geographic location).

In this study, we examined whether data from GISP could be used to develop predictive models that enable personalized empiric treatment of gonorrhea based on a select set of patient characteristics, including sexual behavior and geographic location, in addition to information related to the local prevalence of AMR gonorrhea. We estimated the effectiveness of these personalized treatment recommendations and their ability to reduce the unnecessary use of newer antibiotics where older treatments were effective, compared with standardized treatment guidelines.

Methods

Our dataset includes the record of isolates evaluated for resistance as part of GISP between 2000–2010. Every month during this period, GISP collected isolates from the first 25 men

diagnosed with urethral gonorrhea from 33 sentinel sites across five geographic regions in the US (Table A in [S1 File](#)) [14]. Information on sexual behavior (determined by the sex of the sexual partner) was collected from each individual. These isolates are then tested for susceptibility to ciprofloxacin (CIP), cefixime (CFX), ceftriaxone (CRO), and other antibiotics. In this study, we use the GISP data between 2000–2010 to develop and evaluate models to predict reduced susceptibility to CIP for an individual with gonorrhea. Reduced susceptibility for CIP is defined as presenting either resistance to CIP (a minimum inhibitory concentration of $\geq 1\mu\text{g}/\text{mL}$) or intermediate resistance to CIP (a minimum inhibitory concentration of $0.125\text{--}0.5\mu\text{g}/\text{mL}$) [14]. Our analysis encompasses the period when CIP was one of the recommendations for first-line therapy for gonorrhea, alongside CRO and CFX. CIP was first recommended for the treatment of gonorrhea in 1993, but by 2006, about 13% of gonorrhea isolates evaluated by GISP were resistant to CIP [6]. In April 2007, the CDC stopped recommending CIP for the empiric treatment of gonorrhea. Instead, between 2007 and 2010, either ceftriaxone or cefixime (“CRO/CFX”) was recommended as first-line treatment [6].

Our objective is to examine, retrospectively, the effectiveness of treatment recommendations that are personalized based on the patient’s geographic location and sexual behavior, and information about the local and regional epidemiology of AMR gonorrhea. We also compare the effectiveness of these personalized treatment recommendations and their ability to reduce the unnecessary use of CRO and CFX with treatment guidelines recommended by CDC during this period.

Predictors and outcomes

To develop our predictive models, we considered eight features: 1) the sexual behavior of the individual who submitted a sample, 2) the geographic region in which the sample was taken, 3) the prevalence of resistance to CIP among the isolates collected last year by the surveillance site, 4) the prevalence of resistance to CIP among the isolates collected last year from the geographic region where the surveillance site is located, and 5 and 6) the change in the prevalence of resistance from the previous year in the surveillance site and the geographic region. Given the variation in the annual number of samples collected from each surveillance site, we included two additional binary features to account for the accuracy of estimates for the prevalence of resistance to CIP: 7) whether the prevalence data was calculated on more than 75 observations for that year, and 8) whether the trend data was calculated on more than 75 observations for each year. These features were based on data made available as part of the GISP survey. Other variables of interest, such as age, or ethnicity, were not available as part of this dataset.

Two variables (region and sexual orientation) were nominal categorical variables, and we used one-hot encoding to include them in our analysis (§S1.1 of SI). The geographic location was available for all individuals. For sexual orientation, we coded entries with missing, other, or unknown as “missing” and kept them in the analysis. We standardized all continuous variables (the previous year’s regional and clinical prevalence of CIP resistance and the change in regional and clinical prevalence).

The outcome we wanted to predict was reduced susceptibility to CIP for a gonococcal infection given the patient’s characteristics (e.g., geographic location and sexual orientation). Therefore, all samples with CIP resistances, including samples with additional resistance to other antibiotics, were labelled as positive instances. During our study period (2000–2010), the prevalence of resistance to each of CRO and CFX was $<0.01\%$. Hence, we assume that all gonococcal infections during this period were susceptible to CRO and CFX and could be successfully treated with these antibiotics [5].

Model development

As described above, the guideline for the empiric treatment of all individuals with gonorrhea was changed in 2007 to remove CIP from the recommended treatment guidelines, leaving only CRO or CFX available as recommended treatments. Our goal is to evaluate the utility of personalized treatment recommendations during years before and after the change in recommendation occurred. As such, for each year between 2005–2010, we used data during the prior five years to develop and validate predictive models to inform personalized treatment recommendations during the year of interest, which we refer to as the “target year”. We restricted the training set to just five years as there were large changes in the prevalence of CIP resistance, so using data from more than five years before the target year in question may have reduced the accuracy of our models.

To develop predictive models for each target year, we followed the procedures outlined in the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) [15]. For each year, we considered three supervised Machine Learning (ML) models: a multilayer perceptron (neural network), a logistic regression model, and a random forest. We developed all of our machine learning models using the scikit-learn package in Python [16].

The predictive models we considered here estimate the probability that a gonococcal infection has reduced susceptibility to CIP, given patient characteristics. If this probability is greater than a classification threshold $\lambda(p)$, we classified the infection as “CIP-non-susceptible”; otherwise, the infection is classified as “CIP-susceptible”. Changing the classification threshold, $\lambda(p)$, allows us to change the sensitivity and specificity of each model and quantify the trade-off between the model’s ability to recommend effective personalized treatment versus to prevent the unnecessary prescription of CRO/CFX when the infection was susceptible to CIP.

To improve the predictive power of each year’s models, we performed hyperparameter tuning using a random search algorithm and ten-fold cross-validation, which has been shown to reduce the bias and variance of predictive models [17] (§S1.2 of SI). We then used permutation importance to determine which features increased the area under the Receiver Operating Characteristic curve (auROC). After removing the features that decreased the auROC, we conducted another round of hyperparameter tuning using 10-fold cross-validation for each model in each target year.

Model validation and performance

In line with TRIPOD recommendations [15], we used temporal validation to assess the performance of our models. For each target year, data from the five previous years was used for training the model, hyperparameter tuning, and feature selection. We then used the records collected during the target year to evaluate the model performance based on three criteria: the area under the receiver operating characteristic curve (auROC), the Matthews correlation coefficient (MCC), and the F1 score. As the values of the latter two depend on the classification threshold $\lambda(p)$, we calculated scores over a range of classification thresholds.

To evaluate the sensitivity of our estimates to the observations collected during training years, the selected features, and the selected values for each model’s hyperparameters, we also report 95% bootstrap confidence intervals, calculated using the method described in §S1.3.

We also conducted Leave-One-Out Cross Validation (LOOCV) to evaluate model generalizability beyond surveillance sites included in GISP. For each target year, we first developed models using the data from all surveillance sites excluding data from one of the sites. Then, we temporally validated the models using observations from the excluded site submitted during the target year.

Effectiveness of personalized treatment recommendations and their impact on the unnecessary use of CRO/CFX

For each target year, we report the percentage of individuals that would 1) receive effective treatment (i.e., the treatment that matches the resistance profile of their gonorrhea strain) and 2) be unnecessarily treated with CRO/CFX therapy while their infecting strain is susceptible to CIP. For example, following the standardized guidelines between 2007–2010, which recommended treatment with CRO/CFX, would result in 100% treatment effectiveness but at the cost of unnecessarily using CRO/CFX for $(100(1 - \mu))\%$ for individuals with gonorrhea, where (μ) is the prevalence of resistance to CIP in that year. The additional details on how these measures are calculated are provided in §S1.5 of SI.

These metrics allow us to estimate the change in the treatment effectiveness and the unnecessary use of CRO/CFX if personalized treatments are followed instead of standardized guidelines. We also included a comparison with the count of individuals who would have received effective treatment under the revised CDC guidelines years 2005–2006, which recommended against using CIP for those residing in California and Hawai'i and those identifying as MSM.

Results

Among the isolates collected during 2000–2010, 8.1% had resistance to CIP. The prevalence of resistance to CIP was highest in the West and among the MSM population (Table 1). Models developed for each target year had an auROC greater than 0.5, indicating that they were better than random chance at predicting reduced susceptibility to CIP (Fig 1). The higher the auROC values, the better the model was at distinguishing between ciprofloxacin-susceptible and resistant strains. Models developed for each target year had an auROC greater than 0.5, indicating that they were better than random chance at predicting reduced susceptibility to CIP (Fig 1). The higher the auROC values, the better the model was at distinguishing between ciprofloxacin-susceptible and resistant cases. The neural network and random forest models had comparable performance based on the Matthews correlation coefficients (MCCs) and F1 scores but outperformed the logistic regression models (Fig B in S1 File). Based on these metrics, the neural network model had the highest and most consistent performance scores across target years, indicating that it could respond to changes in the prevalence of CIP-R across time. We, therefore, focus the remainder of our analyses on neural network models, though results for the other ML models are shown in §S2 of SI.

For the neural network models, the most important features included whether the patient was MSW or MSM, the local prevalence of resistance to CIP (“Local CIP-R”), and the change

Table 1. Select demographic information of GISP isolates collected between 2000–2010 (N = 64,445). Overall, for this period, 8.08% of isolates had resistance to CIP.

Variable	% of isolates with resistance to CIP	% of all isolates
Geographic region		
Midwest	1.6	21.2
Northeast	7.7	10.9
Southeast	5.1	18.5
Southwest	4.6	13.9
West Sexual behavior	14.9	35.5
MSM	23.6	17.0
MSW	4.3	76.1
MSMW	17.7	3.6
Unknown/Missing	4.5	3.4

<https://doi.org/10.1371/journal.pdig.0000549.t001>

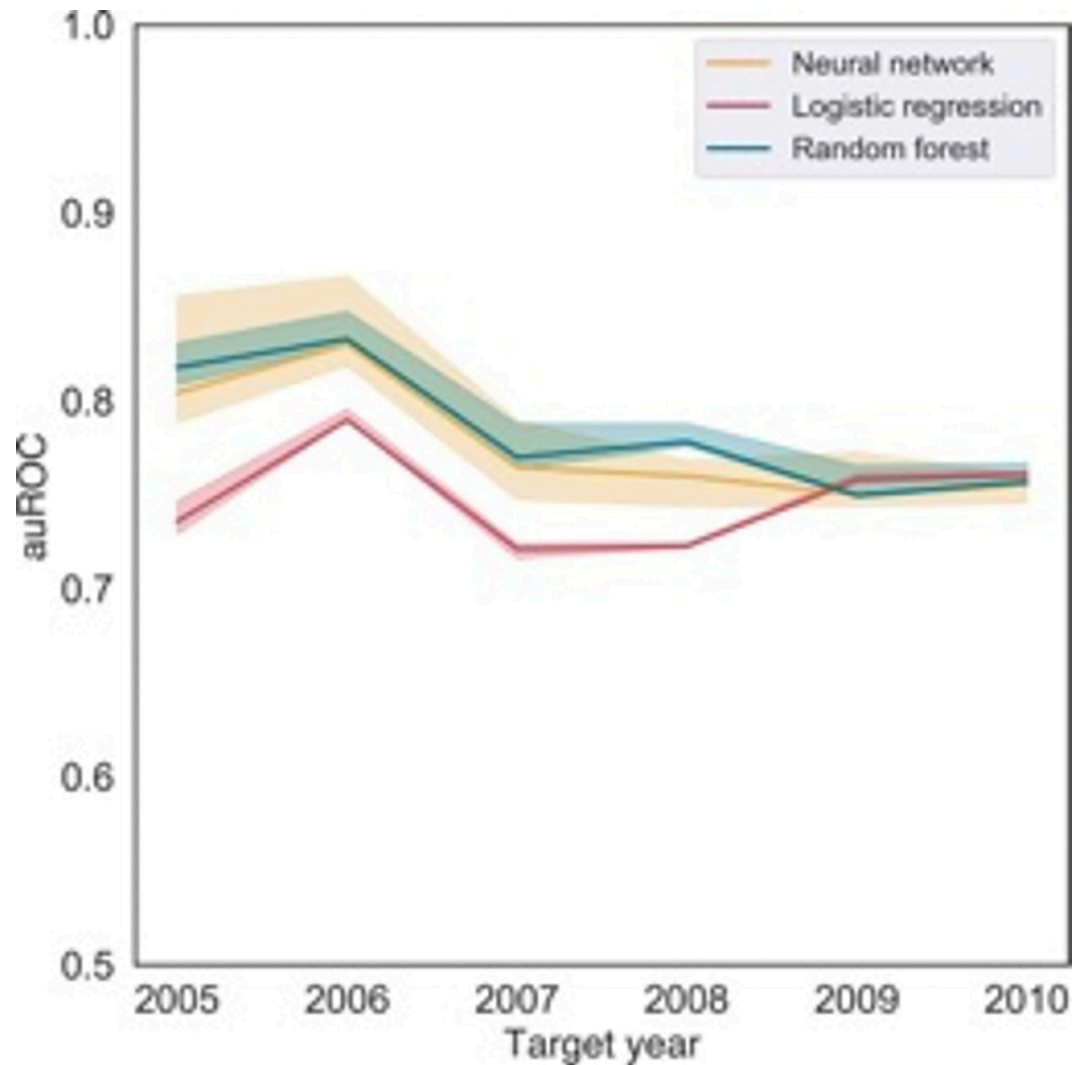


Fig 1. Area under the receiver operating curve (auROC) for each predictive model. For each target year, the model was trained on data from the five previous years and then applied to the data collected during the target year to calculate auROC. The shaded regions show the 95% bootstrap confidence interval, calculated as per the algorithm in [S1 File](#).

<https://doi.org/10.1371/journal.pdig.0000549.g001>

in the local and regional CIP-R prevalence (“Local/Regional change in CIP-R”) ([Fig 2](#)). Other highly important features included the region in which the sample was collected. For each target year, only features that either improved the auROC or did not affect it were included in the final model. Included features and their importance for the logistic regression and random forest models are shown in [Fig B](#) in [S1 File](#).

Under standardized guidelines, the proportion of patients receiving effective treatment or being unnecessarily treated with CRO/CFX depended on the proportion of patients who were prescribed CIP versus CRO/CFX. This is represented by the grey dashed lines in [Fig 3](#); in each panel, the scenario where all patients are treated with CIP (which we refer to as “CIP”) is represented by a star, and the scenario where all patients are treated with CRO or CFX (which we refer to as “CRO/CFX”) is represented by a black box. These lines represent the performance of standardized guidelines when a portion of individuals receive CIP, and the remaining portion receives CRP/CFX.

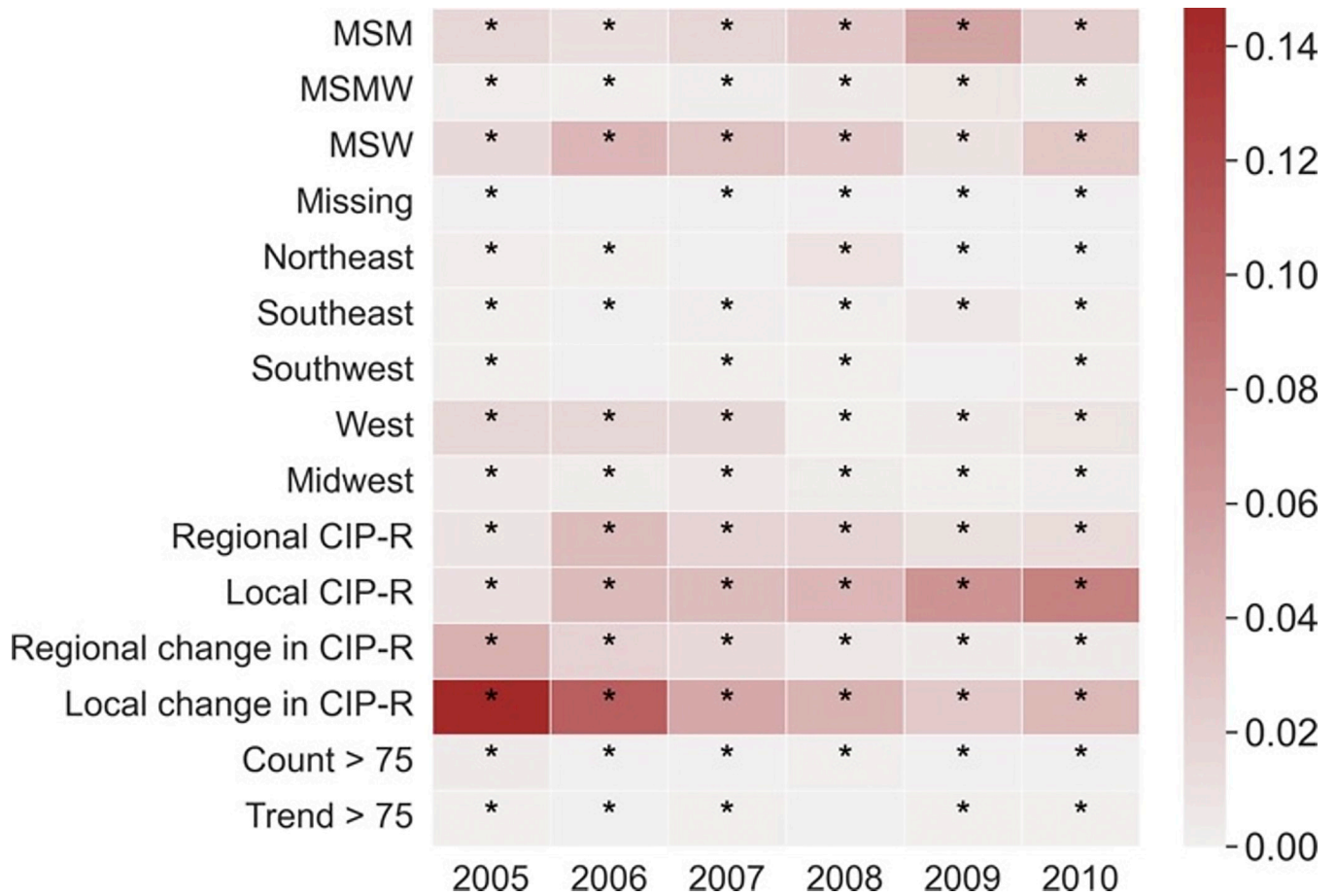


Fig 2. The contribution of features to the area under the receiver operating curve (auROC) score across each model and target year. The color bar represents the contribution of each factor, measured by the feature importance score provided by the permutation importance algorithm. Only features with a positive effect on the training auROC were used in the final models. These features are represented by (*). CIP-R is the prevalence of resistance to ciprofloxacin.

<https://doi.org/10.1371/journal.pdig.0000549.g002>

The “X” marks the CDC’s revised treatment guidelines where the majority of patients were prescribed CIP but patients in Hawai’i, California, and those identify as MSM received CRO or CFX. Compared to scenarios “CIP” or “CRO/CFX”, these revised CDC guidelines improved the percentage of individuals receiving effective treatment but also increased the unnecessary use of CRO/CFX.

For each target year, for a given threshold for treatment effectiveness (say, 98%, which is represented by vertical black dotted lines in Fig 3), following the personalized treatment provided by neural network models could have reduced the unnecessary use of CRO/CFX. The magnitude of this reduction is measured by the vertical distance between the dotted line and the bold curve in each panel of Fig 3 and is dependent on the model’s classification threshold (*p*), represented by numbers on curves in Fig 3. For each target year during 2005–2010, we could identify a classification threshold such that the effectiveness of personalized treatment regimens is maintained at 98%. At this effectiveness threshold, following the personalized recommendations provided by the neural network model could have reduced the unnecessary use of CRO/CFX by an average of 32.8 percentage points compared to recommendations provided by standardized guidelines during 2005–2010.

The results of Leave-One-Out Cross Validation (LOOCV) suggest that the ability of predictive models to reduce the unnecessary use of CRO/CFX while maintaining a high level of

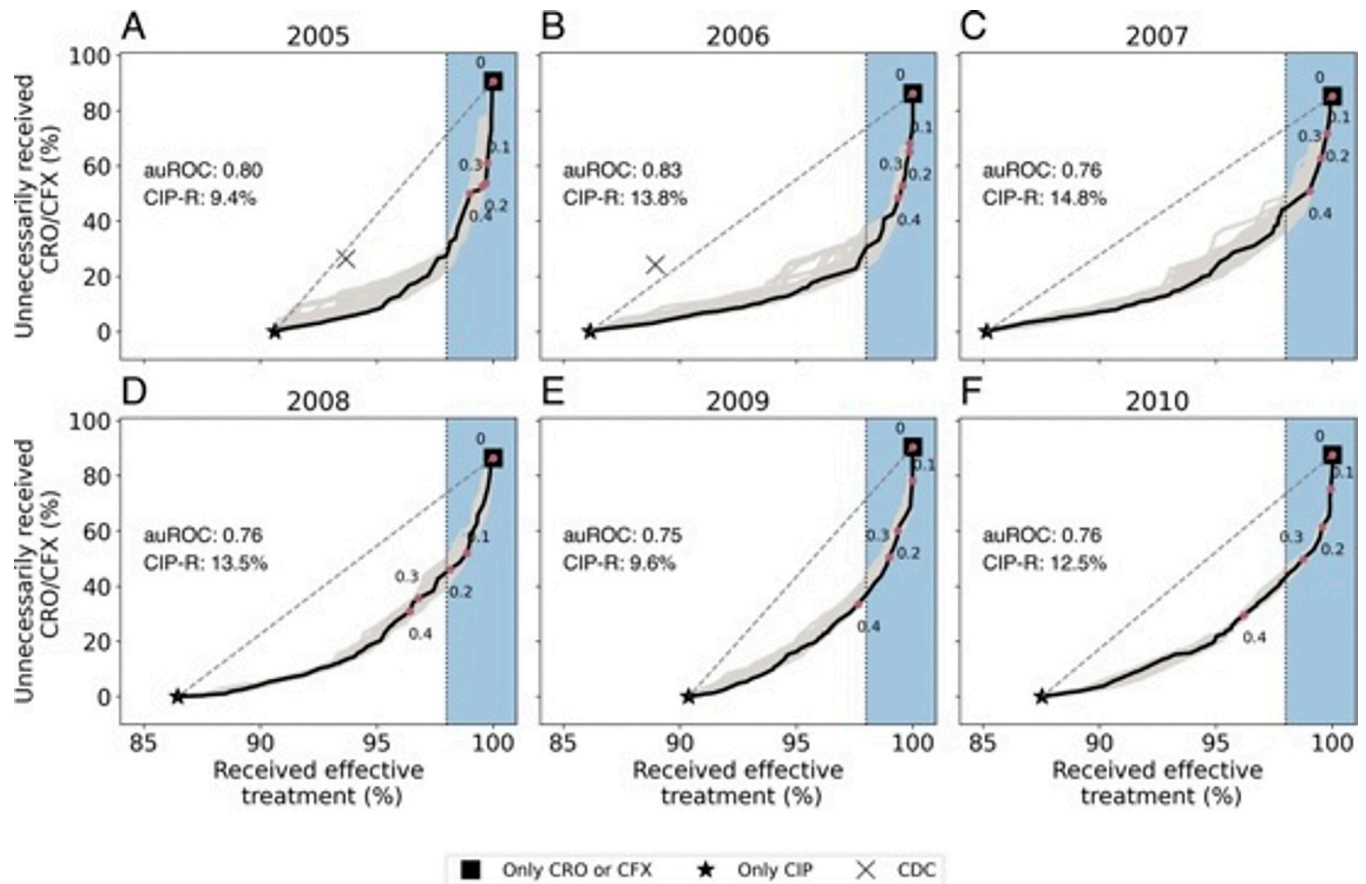


Fig 3. The proportion of patients with gonorrhea receiving effective treatment or being unnecessarily treated with CRO/CFX when CIP could have been effective. Numbers on the curves represent different classification thresholds. Higher classification thresholds decrease the proportion receiving effective treatment and unnecessarily treated with ceftriaxone or cefixime (CRO/CFX). The grey dashed line shows the impact if a random portion of individuals with gonorrhea receive CIP and the rest receive CRO/CFX therapy. The “X” denotes the scenario where no one in California, Hawai’i, or who identified as men who had sex with men (MSM) received ciprofloxacin (CIP) as recommended by CDC treatment guidelines in 2005 and 2006. The blue regions represent predictive models that would result in 98% treatment effectiveness. The grey curves show the 95% bootstrap region generated using the bootstrapping procedure described in S1.3.

<https://doi.org/10.1371/journal.pdig.0000549.g003>

treatment effectiveness varies across different locations (Fig 4). At low classification thresholds, the proportion receiving effective treatment can be maintained at > 95% while reducing the unnecessary prescription of CRO/CFX by between 4 and 100 percentage points, in some cases, therefore, eliminating the unnecessary prescription of CRO/CFX. However, for regions with higher than the national average CIP-R prevalence (i.e., pink curves in Fig 4), the models trained on national data would have been less effective in reducing the unnecessary use of CRO/CFX while maintaining the treatment effectiveness at 95%.

Discussion

The common methods to diagnose gonorrhea do not provide information about the susceptibility of the infecting strain to available antibiotics [4,18]. The treatment of gonorrhea, therefore, remains empiric and according to standardized guidelines, which are determined at the national level. Although these standardized guidelines are easy to communicate, they may result in prescribing an ineffective therapy and diminishing the effective lifespan of existing antibiotics. In this retrospective study, we used data from GISP collected between 2000–2010

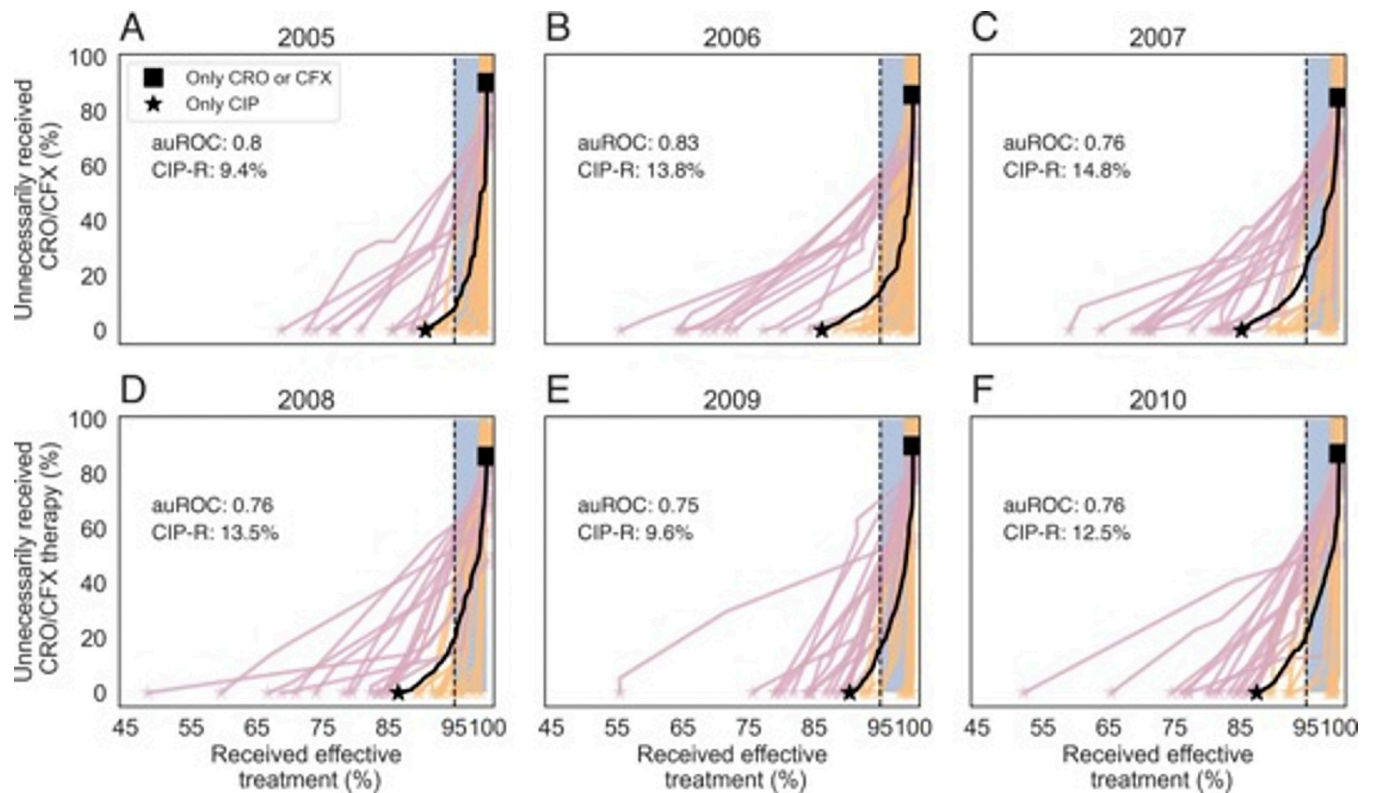


Fig 4. Leave-One-Out Cross Validation (LOOCV). Each individual curve corresponds to the surveillance site the model was tested on, but the data from this surveillance site was excluded when developing the model. Pink lines are clinics where the resistance to ciprofloxacin (CIP-R) prevalence was above the national average, and orange lines are where it was below. The black curve corresponds to the model that is developed using data from all surveillance sites during the five years prior to the target year. The blue regions represent predictive models that would result in 95% treatment effectiveness.

<https://doi.org/10.1371/journal.pdig.0000549.g004>

to develop and evaluate machine learning (ML) models that use basic demographic information to identify reduced susceptibility to CIP, one of the recommended first-line antibiotics in the US until 2007 [6]. We estimated the impact of following the recommendations provided by these predictive models on the probability of using effective antibiotics and the unnecessary use of either ceftriaxone or cefixime (CRO/CFX), which were the recommended first-line therapies between 2007–2010 [6].

Our analysis showed that following the personalized recommendation provided by these ML models could have improved the probability of receiving effective treatment and reduced the unnecessary use of CRO/CFX among patients with gonorrhea between 2005–2010 in the US. In the neural network and random forest models, we found that regional and local epidemiological factors were generally the most important feature, alongside sexual behavior (Fig 2 and Fig B in S1 File). These features are consistent with previous studies suggesting that CIP resistance is more common amongst some “core” groups defined by sexual behavior [14,19,20], and that the local prevalence of resistance is an important factor in determining empiric treatment guidelines [14,21]. However, in both models, the majority of features were considered important (having a positive impact on the auROC). Considerably fewer features were included in the logistic regression model, though important features included sexual behavior and the region in which the sample was collected.

Of the three models developed (logistic regression, neural network, and random forest), the neural network model had the highest and most stable auROC (Fig 1), MCC and F1 scores across all test years (SS2 in SI). We, therefore, focus the majority of our analyses on the results

of this model. The ability of the models considered here to improve treatment effectiveness and reduce the unnecessary use of newer antibiotics depended on the selected classification threshold. Across all classification thresholds, there was a trade-off between the proportion of patients receiving effective treatment and the unnecessary use of CRO/CFX (Fig 3 and Figs F-G in S1 File). Yet all models could have led to significant reductions in the unnecessary use of CRO/CFX while maintaining the efficacy of first-line treatment above 98%.

Between 2000 and 2007, the CDC revised its empiric treatment guidelines in response to an increasing prevalence of CIP-resistant infections. Whilst these revised guidelines improved the proportion of patients receiving effective treatment, they also increased the proportion that were unnecessarily treated with CRO/CFX when CIP would have been effective. The approach that we proposed here provides a more systematic method to translate data collected through GISP into personalized treatment recommendations, reducing the unnecessary use of new antibiotics without affecting the proportions of receiving effective treatment.

The optimal choice of classification threshold requires quantifying the positive and negative consequences of using predictive models with imperfect sensitivity and specificity (Figs C-E in S1 File). The acceptable reduction in treatment efficacy for a given reduction in the unnecessary use of CRO/CFX may vary between decision-makers. This could be more carefully investigated through a cost-effectiveness analysis.

Our study has several limitations. First, our dataset is limited only to the surveillance sites included in GISP, and hence the performance of predictive models developed here for locations not included in GISP remains unknown. However, using LOOCV, we found that when the model was used to make predictions for clinics not included in the training set, treatment efficacy could remain > 95% while reducing—and in some cases eliminating—the need for CRO/CFX. Our LOOCV analysis suggests that the model's predictions were worse for clinics where the CIP-R was higher than the national average. Second, implicit in our analysis is that equal weighting is given to ensuring treatment efficacy while also preventing the unnecessary use of CRO/CFX. However, in practice, a greater priority might be given to ensuring that patients receive an antibiotic that matches the susceptibility profile of their infection, causes the minimal side-effects and is inexpensive. This trade-off could be investigated through cost-effective analyses accounting for the consequences of prescribing ineffective treatment and any benefits gained from extending the clinical usefulness of existing antibiotics. Third, we included a limited number of features in our model (namely, the sexual behavior of the patient, geographic region, the prevalence of CIP resistance, and the change in regional prevalence of CIP resistance). However, other socio-demographic and health factors are believed to be associated with having a drug-resistant gonorrhoea infection [13,19,22–28]. Including these variables may enhance our model's predictive power and could be considered in future analyses. Nevertheless, despite the limited range of features, our ML models displayed reasonable predictive capabilities and demonstrated the potential to reduce the unnecessary use of CRO/CFX. Moreover, the selected features can be easily obtained at the point of care, facilitating their implementation in clinical settings. However, we do acknowledge that knowing the features which are important to the models' prediction does not indicate how those features are being used by the model. The explainability of machine learning models is an area of ongoing and much-needed research to facilitate the use of machine learning models in practice [29,30]. Fourth, between 2000–2010, GISP included isolates only from the first 25 men at each surveillance site presenting with urethral gonorrhoea. However, considering the potential variance in resistance dissemination among population groups, the performance of our ML models developed in this study may differ for other sexual behavior populations or individuals infected at anatomical sites beyond the urethra [13,19,31,32]. It is worth noting that GISP was expanded in 2017 to encompass isolates from multiple anatomical sites and from men and women

[18,33]. The incorporation of data from the enhanced Gonococcal Isolate Surveillance Project (eGISP) could mitigate this limitation in future analyses. Since 2020, the CDC has recommended single-drug therapy with ceftriaxone for the empiric treatment of gonorrhea [8]. As the prevalence of resistance to ceftriaxone is still extremely low (<0.002% among GISP isolates collected between 2000–2019), it would not be possible to develop models to predict the resistance to ceftriaxone given a patient's characteristics. However, the ML methods we described here could be used to guide the use of ceftriaxone in the empiric treatment of gonorrhea in the future when the resistance to this antibiotic starts to spread more widely. As suggested by our analysis, we anticipate that the treatment recommendation provided by ML models outperforms the ad-hoc and heuristic refinement of guidelines, as occurred during 2000–2007. Despite the global rise in AMR infections, the treatment of many bacterial infections (such as tuberculosis and gonorrhea) remains empiric due to a lack of access to drug-susceptibility testing (DST) or delays in receiving the results of DSTs. Our findings contribute to the emerging evidence that ML models trained on data from national AMR surveillance systems could identify resistance based on the patient's demographic and clinical information[34–40]. We showed that using predictive models to guide personalized treatment recommendations could reduce the use of newer antibiotics, and potentially extending their useful clinical lifespan. Future studies could investigate the potential of these ML models to inform the selection of antibiotics, at the point-of-care, for patients with AMR infections.

Supporting information

S1 File. S1: Supplementary methods. S1.1 GISP data and data pre-processing. S1.2 Hyperparameter tuning. S1.3 Bootstrap algorithm to calculate confidence intervals for performance measures. S1.4 Model specificity and sensitivity. S1.5 Calculating the effectiveness of personalized treatment recommendations and their impact on the unnecessary use of CRO/CFX. **S2: Supplementary results.** S2.1 Permutation importance and feature selection. S2.2 Model performance across target years. S2.3 Percentage receiving effective treatment and unnecessary CRO/CFX for logistic regression and random forest. S2.4 Leave-one-out cross-validation for logistic regression and random forest (DOCX)

Acknowledgments

The authors would like to acknowledge CDC's Gonococcal Isolate Surveillance Project for the data used in these analyses. Additionally, the authors would like to thank the participating GISP sentinel sites and GISP regional laboratories for their contributions to GISP and collection of project data.

Author Contributions

Conceptualization: Rachel E. Murray-Watson, Yonatan H. Grad, Reza Yaesoubi.

Formal analysis: Rachel E. Murray-Watson.

Funding acquisition: Reza Yaesoubi.

Investigation: Rachel E. Murray-Watson.

Methodology: Rachel E. Murray-Watson, Reza Yaesoubi.

Project administration: Reza Yaesoubi.

Resources: Reza Yaesoubi.

Software: Rachel E. Murray-Watson.

Supervision: Yonatan H. Grad, Reza Yaesoubi.

Validation: Rachel E. Murray-Watson, Reza Yaesoubi.

Visualization: Rachel E. Murray-Watson.

Writing – original draft: Rachel E. Murray-Watson, Yonatan H. Grad, Reza Yaesoubi.

Writing – review & editing: Rachel E. Murray-Watson, Yonatan H. Grad, Sancta B. St. Cyr, Reza Yaesoubi.

References

1. Yang F. and Yan J. Antibiotic resistance and treatment options for multidrug-resistant gonorrhoea. *Infectious Microbes & Diseases*, 2(2):67–76, 2020.
2. Centers for Disease Control and Prevention (CDC). COVID19: US impact on Antimicrobial Resistance, Special Report 2022. Atlanta, GA: US Department of Health and Human Services. Available online: <https://www.cdc.gov/drugresistance/covid19>. (Accessed: 11th April 2022), 2022.
3. CDC et al. Antibiotic resistance threats in the United States, 2019. US Department of Health and Human Services, Centers for Disease Control, 2019.
4. Schlanger K., Learner E. R., Pham C. D., Mauk K., Golden M., Wendel K. A., Amsterdam L., McNeil C. J., Johnson K., Nguyen T. Q., et al. Strengthening the US response to resistant gonorrhoea: An overview of a multisite program to enhance local response capacity for antibiotic-resistant *Neisseria gonorrhoeae*. *Sexually Transmitted Diseases*, 48(12):S97–S103, 2021. <https://doi.org/10.1097/OLQ.0000000000001545> PMID: 34475362
5. Centers for Disease Control and Prevention (CDC). CDC—Gonococcal Isolate Surveillance Project (GISP). <https://www.cdc.gov/std/gisp/default.htm>, 2023. (Accessed: 15th January 2023).
6. CDC et al. Update to CDC’s sexually transmitted diseases treatment guidelines, 2006: fluoroquinolones no longer recommended for treatment of gonococcal infections. *Morbidity and Mortality Weekly Report*, 56(14):332–336, 2007. PMID: 17431378
7. Blank S. and Daskalakis D. C. *Neisseria gonorrhoeae*-rising infection rates, dwindling treatment options. *The New England Journal of Medicine*, 379(19):1795–1797, 2018. <https://doi.org/10.1056/NEJMp1812269> PMID: 30403946
8. St Cyr S., Barbee L., Workowski K. A., Bachmann L. H., Pham C., Schlanger K., Torrone E., Weinstock H., Kersh E. N., and Thorpe P. Update to CDC’s treatment guidelines for gonococcal infection, 2020. *Morbidity and Mortality Weekly Report*, 69(50):1911, 2020. <https://doi.org/10.15585/mmwr.mm6950a6> PMID: 33332296
9. Fairley C. K., Hocking J. S., Zhang L., and Chow E. P. Frequent transmission of gonorrhoea in men who have sex with men. *Emerging Infectious Diseases*, 23(1):102, 2017. <https://doi.org/10.3201/eid2301.161205> PMID: 27983487
10. Stefanelli P., Vescio M. F., Landini M. P., Dal Conte I., Matteelli A., Cristaudo A., Gaino M., Cusini M., Barbui A. M., Mencacci A., et al. Time trend analysis (2009–2016) of antimicrobial susceptibility in *Neisseria gonorrhoeae* isolated in Italy following the introduction of the combined antimicrobial therapy. *PLoS One*, 12(12):e0189484, 2017. <https://doi.org/10.1371/journal.pone.0189484> PMID: 29240786
11. Unemo M., Seifert H. S., Hook E. W. III, Hawkes S., Ndowa F., and Dillon J. R. Gonorrhoea. *Nature Reviews Disease Primers*, 5(1):79, 2019. <https://doi.org/10.1038/s41572-019-0128-6> PMID: 31754194
12. CDC. Sexually Transmitted Disease Surveillance 2020: Gonococcal Isolate Surveillance Project National Profile. 2022.
13. Visser M., Götz H. M., van Dam A. P., and van Benthem B. H. Trends and regional variations of gonococcal antimicrobial resistance in the Netherlands, 2013 to 2019. *Eurosurveillance*, 27(34):2200081, 2022. <https://doi.org/10.2807/1560-7917.ES.2022.27.34.2200081> PMID: 36017715
14. CDC. Sexually Transmitted Disease Surveillance 2019: Gonococcal Isolate Surveillance Project (GISP) Supplement and Profiles. 2021.
15. Moons K. G. M., Altman D. G., Reitsma J. B., Ioannidis J. P. A., Macaskill P., Steyerberg E. W., Vickers A. J., Ransohoff D. F., and Collins G. S. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of Internal Medicine*, 162(1):W1–W73, 2015. <https://doi.org/10.7326/M14-0698> PMID: 25560730

16. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., and Duchesnay E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
17. James G., Witten D., Hastie T., Tibshirani R., et al. *An introduction to statistical learning*, volume 112. Springer, 2013.
18. Kersh E. N., Pham C. D., Papp J. R., Myers R., Steece R., Kubin G., Gautom R., Nash E. E., Sharpe S., Gernert K. M., et al. Expanding US laboratory capacity for *Neisseria gonorrhoeae* antimicrobial susceptibility testing and whole-genome sequencing through the CDC's Antibiotic Resistance Laboratory Network. *Journal of Clinical Microbiology*, 58(4):e01461–19, 2020.
19. Abraha M., Egli-Gany D., and Low N. Epidemiological, behavioural, and clinical factors associated with antimicrobial-resistant gonorrhoea: a review. *F1000Research*, 7, 2018. <https://doi.org/10.12688/f1000research.13600.1> PMID: 29636908
20. Fuertes de Vega I., Baliu-Piqué C., Bosch Mestres J., Vergara Gómez A., Vallés X., and Alsina Gibert M. Risk factors for antimicrobial-resistant *Neisseria gonorrhoeae* and characteristics of patients infected with gonorrhoea. *Enfermedades Infecciosas y Microbiología Clínica*, 36(3):165–168, 2018. <https://doi.org/10.1016/j.eimc.2016.11.012> PMID: 28094065
21. Yaesoubi R., Cohen T., Hsu K., Gift T. L., Chesson H., Salomon J. A., and Grad Y. H. Adaptive guidelines for the treatment of gonorrhoea to increase the effective life span of antibiotics among men who have sex with men in the United States: A mathematical modeling study. *PLoS Medicine*, 17(4): e1003077, 2020. <https://doi.org/10.1371/journal.pmed.1003077> PMID: 32243443
22. Bernstein K. T., Marcus J. L., Nieri G., Philip S. S., and Klausner J. D. Rectal gonorrhoea and chlamydia reinfection is associated with increased risk of HIV seroconversion. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 53(4):537–543, 2010. <https://doi.org/10.1097/QAI.0b013e3181c3ef29> PMID: 19935075
23. Pradier L. and Bedhomme S. Ecology, more than antibiotics consumption, is the major predictor for the global distribution of aminoglycoside-modifying enzymes. *eLife*, 12:e77015, 2023. <https://doi.org/10.7554/eLife.77015> PMID: 36785930
24. Kirkcaldy R. D., Harvey A., Papp J. R., Del Rio C., Soge O. O., Holmes K. K., Hook E. W. III, Kubin G., Riedel S., Zenilman J., et al. *Neisseria gonorrhoeae* antimicrobial susceptibility surveillance—the gonococcal isolate surveillance project, 27 sites, United States, 2014. *Morbidity and Mortality Weekly Report: Surveillance Summaries*, 65(7):1–19, 2016. <https://doi.org/10.15585/mmwr.ss6507a1> PMID: 27414503
25. Quilter L. A. S., St Cyr S. B., Hong J., Asbel L., Bautista I., Carter B., Casimir Y., Denny M., Ervin M., Gomez R., et al. Antimicrobial susceptibility of urogenital and extragenital *Neisseria gonorrhoeae* isolates among men who have sex with men: Strengthening the US Response to Resistant Gonorrhoea and Enhanced Gonococcal Isolate Surveillance Project, 2018 to 2019. *Sexually Transmitted Diseases*, 48(12S):S111–S117, 2021. <https://doi.org/10.1097/OLQ.0000000000001548> PMID: 34475363
26. Wang S. A., Harvey A. B., Conner S. M., Zaidi A. A., Knapp J. S., Whittington W. L. H., Del Rio C., Judson F. N., and Holmes K. K. Antimicrobial resistance for *Neisseria gonorrhoeae* in the United States, 1988 to 2003: the spread of fluoroquinolone resistance. *Annals of Internal Medicine*, 147(2):81 <https://doi.org/10.7326/0003-4819-147-2-200707170-00006> PMID: 17638718
27. Centers for Disease Control, Prevention (CDC), et al. Cephalosporin susceptibility among *Neisseria gonorrhoeae* isolates—United States, 2000–2010. *MMWR. Morbidity and Mortality Weekly Report*, 60(26):873–877, 2011.
28. Kirkcaldy Robert D., Zaidi Akbar, Hook Edward W. III, Holmes King K., Soge Olusegun, Carlos del Rio Geraldine Hall, Papp John, Bolan Gail, and Weinstock Hillard S. *Neisseria gonorrhoeae* antimicrobial resistance among men who have sex with men and men who have sex exclusively with women: the gonococcal isolate surveillance project, 2005–2010. *Annals of Internal Medicine*, 158(5_Part_1):321–328, 2013. <https://doi.org/10.7326/0003-4819-158-5-201303050-00004> PMID: 23460055
29. Sana Tonekaboni, Shalmali Joshi, Melissa D. McCradden, and Anna Goldenberg. What clinicians want: Contextualizing explainable machine learning for clinical end use. In *Finale Doshi-Velez, Jim Fackler, Ken Jung, David Kale, Rajesh Ranganath, Byron Wallace, and Jenna Wiens, editors, Proceedings of the 4th Machine Learning for Healthcare Conference*, volume 106 of *Proceedings of Machine Learning Research*, pages 359–380. PMLR, 09–10 Aug 2019.
30. Belle Vaishak and Papantonis Ioannis. Principles and practice of explainable machine learning. *Frontiers in Big Data*, 4:688969, 2021. <https://doi.org/10.3389/fdata.2021.688969> PMID: 34278297
31. Furuya Ryusaburo, Onoye Yasuhiko, Kanayama Akiko, Saika Takeshi, Iyoda Takako, Tatewaki Mitsuhiro, Matsuzaki Kaoru, Kobayashi Intetsu, and Tanaka Masatoshi. Antimicrobial resistance in clinical isolates of *Neisseria subflava* from the oral cavities of a Japanese population. *Journal of Infection and Chemotherapy*, 13:302–304, 2007. <https://doi.org/10.1007/s10156-007-0541-8> PMID: 17982718

32. Laumen Jolein G. E., Christophe Van Dijck Said Abdellati, Sheeba S. Manoharan-Basil, Irith De Baetselier, Delphine Martiny, Tania Crucitti, and Chris Kenyon. Markedly reduced azithromycin and ceftriaxone susceptibility in commensal *Neisseria* species in clinical samples from Belgian men who have sex with men. *Clinical Infectious Diseases*, 72(2):363–364, 2021.
33. Sancta St Cyr, Kristen Kreisel, and Pham Cau. Gonococcal Isolate Surveillance Project (GISP) and Enhanced GISP (eGISP) protocol. 2022.
34. You Shiyang, Chitwood Melanie H., Gunasekera Kenneth S., Crudu Valeriu, Codreanu Alexandru, Ciobanu Nelly, Furin Jennifer, Cohen Ted, Warren Joshua L., and Yaesoubi Reza. Predicting resistance to fluoroquinolones among patients with rifampicin-resistant tuberculosis using machine learning methods. *PLOS Digital Health*, 1(6):e0000059, 2022. <https://doi.org/10.1371/journal.pdig.0000059> PMID: 36177394
35. Feretzakis Georgios, Loupelis Evangelos, Sakagianni Aikaterini, Kalles Dimitris, Martsoukou Maria, Lada Malvina, Skarmoutsou Nikoleta, Christopoulos Constantinos, Valakis Konstantinos, Velentza Aikaterini, et al. Using machine learning techniques to aid empirical antibiotic therapy decisions in the intensive care unit of a general hospital in Greece. *Antibiotics*, 9(2):50, 2020. <https://doi.org/10.3390/antibiotics9020050> PMID: 32023854
36. Ohad Lewin-Epstein Shoham Baruch, Hadany Lilach, Stein Gideon Y., and Obolski Uri. Predicting antibiotic resistance in hospitalized patients by applying machine learning to electronic medical records. *Clinical Infectious Diseases*, 72(11):e848–e855, 2021. <https://doi.org/10.1093/cid/ciaa1576> PMID: 33070171
37. Mintz Igor, Chowers Michal, and Obolski Uri. Prediction of ciprofloxacin resistance in hospitalized patients using machine learning. *Communications Medicine*, 3(1):43, 2023. <https://doi.org/10.1038/s43856-023-00275-z> PMID: 36977789
38. Stracy Mathew, Snitser Olga, Yelin Idan, Amer Yara, Parizade Miriam, Katz Rachel, Rimler Galit, Wolf Tamar, Herzel Esma, Koren Gideon, et al. Minimizing treatment-induced emergence of antibiotic resistance in bacterial infections. *Science*, 375(6583):889–894, 2022. <https://doi.org/10.1126/science.abg9868> PMID: 35201862
39. Oonsivilai Mathupanee, Mo Yin, Luangasanatip Nantasit, Lubell Yoel, Miliya Thyl, Tan Pisey, Loeuk Lorn, Turner Paul, and Cooper Ben S. Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children’s hospital in Cambodia. *Wellcome Open Research*, 3, 2018.
40. Feretzakis Georgios, Sakagianni Aikaterini, Loupelis Evangelos, Kalles Dimitris, Skarmoutsou Nikoleta, Martsoukou Maria, Christopoulos Constantinos, Lada Malvina, Petropoulou Stavroula, Velentza Aikaterini, et al. Machine learning for antibiotic resistance prediction: a prototype using off-the-shelf techniques and entry-level data to guide empiric antimicrobial therapy. *Healthcare Informatics Research*, 27(3):214–221, 2021. <https://doi.org/10.4258/hir.2021.27.3.214> PMID: 34384203