medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Stable temporal relationships as a first step towards causal inference: an application to antibiotic resistance

Avi Baraz[1,2,3], Michal Chowers[4,5], Daniel Nevo[3,†], Uri Obolski[1,2,†]

1 School of Public Health, Tel Aviv University, Tel Aviv, Israel

2 Porter School of the Environment and Earth Sciences, Tel Aviv University, Tel Aviv, Israel

3 Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel

4 Meir Medical Centre, Kfar Saba, Israel

5 Sackler Faculty of Medicine, Tel Aviv University, Israel

† These authors contributed equally

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Epidemiological studies often analyze data as static, essentially averaging observed associations across time. Overlooking time trends is especially problematic in settings subject to rapid changes. A prominent example for such a setting is antibiotic resistance, which has reached concerning levels, and poses a global healthcare challenge. Bacteria constantly evolve and hence antibiotic resistance is characterized by time-varying relationships with clinical and demographic covariates. In this paper, we speculate that covariates with a causal effect are expected to have stable relationships with resistance over calendar time. To this end, we applied time-varying coefficient models to a large clinical dataset from an Israeli hospital, and have shown their advantages in describing covariate-resistance relationships. We found both time-stable and time-varying covariate-resistance relationships. These results serve as initial evidence towards causal interpretation of these relationships, as one may expect time-stable rather than time-varying relationships to correspond with causal effects. We further conducted data-driven simulations, that have illustrated how results from time-varying coefficient models must be carefully interpreted with respect to causal claims. Potentially, identification of causal covariate-resistance relationships can lead to new medical interventions and healthcare policies, and improve the generalization of existing predictive models for antibiotic resistance.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Antibiotic resistance is a major global healthcare concern. According to a 2019 report by the Centers for Disease Control and Prevention, in the US alone at least 2.8 million people suffer from antibiotic resistant bacterial infections each year, of which approximately 35,000 result in death (1). Antibiotic use and over-prescription have led to high global resistance levels, prompting the World Health Organization to declare antibiotic resistance a global health crisis, which also threatens the global economy (1,2).

Antibiotic resistant infections emerge through complex biological processes, shaped by evolutionary forces and bacterial population dynamics (3,4). The interplay between these processes and epidemiological-level covariates is still poorly understood. Moreover, estimates of epidemiological-level effects are likely to vary with time, due to the rapid dynamics of bacterial populations in response to environmental changes. For example, resistance frequencies are constantly changing as a result of selective pressure of antibiotic use in medicine (5–8) or agriculture (8–11), demographic changes such as urbanization (12,13) and healthcare-related policies (14–16).

The causes of antibiotic resistance have been the focal point of much research. Randomized controlled trials, which can help unravel these causes, are in many cases difficult and impractical to perform in the context of antibiotic resistance. Hence, many of the studies that attempted to identify the causes of antibiotic resistance were observational. However, observational studies are subject to confounding, which limits their ability to determine if a risk factor causes antibiotic resistance or is only associated with it. Nonetheless, there have been some attempts, including by our group, to overcome confounding from observational studies to estimate the effect of antibiotic use on resistance (e.g. (17–20)).

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In this work, we showcased that estimates of time-stable relationships (over calendar time) between clinical and demographic covariates and resistance, can provide evidence towards causality. If a covariate has a causal effect (or no effect at all) on an outcome variable, one may expect the relationship between them to be time-stable rather than time-varying. The estimated relationship estimate is not necessarily an estimate of the causal effect of the covariate on resistance, but rather evidence towards the existence of a causal relationship. We emphasize that this paper is concerned with calendar time-varying mechanisms resulting from the dynamic nature of infectious diseases. It is not concerned with time-varying treatments and confounders within an individual.

Demonstrating these claims empirically, we estimated temporal relationships between hospitalized patients' bacterial antibiotic susceptibility test results and their corresponding electronic medical records during 2016-19. We modelled these temporal relationships by logistic time-varying coefficient models. Such models are a private case of Generalized Additive Models (GAM) (21,22), that allows flexible modelling of non-linear covariate-outcome relationships in time. Previous research used GAMs in the context of antibiotic resistance. Two such papers dealt with the relationship between population-level antibiotic use and the emergence of resistance (23,24), while another paper utilized GAMs, but not time-varying coefficient models, to model non-linear relationships between hospitalization duration and resistance (25).

We compared our results to standard time-fixed logistic regression models, and explored the differences between them. Our results presented significant time trends in the relationships between risk factors and resistance otherwise overlooked by standard models. Temporal relationships between risk factors and resistance that were estimated as stable, coincided with prior knowledge deeming them causal. On the other hand, temporal relationships estimated as time-varying corresponded to prior knowledge suggesting they were not causal. Finally, we

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

conducted data-driven simulations that demonstrated the need for careful interpretation of obtained time-varying coefficients, as various plausible causal scenarios may give rise to observed time trends.

Methods

The motivating data

The dataset used in this paper includes electronic records of all patients who had a positive bacterial culture between 2016-19 in Meir hospital, Kfar Saba, Israel. The data originated from two types of records: Bacterial antibiotic susceptibility test results, and their corresponding patients' demographic and clinical data. Table 1 presents summary statistics of key patient covariates, stratified by resistance test result (susceptible/resistant), and Tables S1 and S2 describe in detail and present summary statistics of all the covariates used in our analysis.

	Susceptible	**Resistant**
N	9820	1772
Age (mean (SD))	69.63 (20.41)	70.64 (17.43)
# Hospitalizations past year (mean (SD))	2.06 (1.53)	2.62 (1.97)
Male (N (%))	4462 (45.4)	950 (53.6)
Used antibiotics (N (%))	4999 (50.9)	1109 (62.6)
Had MRSA (N (%))	511 (5.2)	220 (12.4)
Culture drawn at the internal unit (N (%))	4725 (48.1)	921 (52.0)
Culture drawn at ER (N (%))	1019 (10.4)	157 (8.9)
Had catheter (N (%))	2436 (24.8)	689 (38.9)
Nosocomial (N (%))	3752 (38.2)	785 (44.3)

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Arrived from an institution (N (%))	1279 (13)	398 (22.5)
Sample location: urine (N (%))	4475 (45.6)	879 (49.6)
Genus: Escherichia (N (%))	4430 (45.1)	698 (39.4)

Table 1. Summary of key covariates used in this study, stratified by gentamicin resistance result. #: number of. MRSA: Methicillin-resistant *Staphylococcus aureus*. ER: emergency room. *Used antibiotics*: hospital use of antibiotics inducing gentamicin resistance during the 6 months prior to the bacterial culture's drawing.

Our analysis focused on resistance to gentamicin. Susceptibility testing to gentamicin was commonly performed, yielding a resistance rate of 15.6%, thus providing us a large sample with satisfactory event rate. Seven bacterial species commonly isolated in hospitalized patients *(Eescherichia coli, Pseudomonas aeruginosa, Klebsiella pneumoniae, Proteus mirabilis, Enterobacter cloacae, Morganella morganii and Citrobacter koseri)* were chosen for the analysis, since each had at least 200 resistance results to gentamicin. To account for the heterogeneity of the species, a categorical covariate indicating the bacteria's genus was incorporated in all our models. In total, 11,592 gentamicin resistance results from the above-mentioned bacterial species were sampled between 2016-19 and included in our analysis.

Statistical methods

Temporal relationships between patients' clinical and demographic covariates, and resistance, were modelled by logistic time-varying coefficient models, which are an extension of standard logistic regression models (21,22). Unlike standard logistic regression models, time-varying coefficient models allow for variation of covariate-outcome relationships across time by replacing some of the time-fixed coefficients with non-parametric functions of time. These models take the form

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

$$logit[p_t(Y_{it} = 1|\boldsymbol{X}_{it})] = \beta_0 + \sum_{m=1}^{l} \beta_m X_{mit} + \sum_{j=l+1}^{p} \beta_j(t) X_{jit},$$

where Y_{it} is the gentamicin resistance, and $\boldsymbol{X}_{it} = (X_{1it}, \ldots, X_{pit})$ is the vector of p covariates of the ith observation, both measured at time t, which is the time when the bacterial culture was drawn from the patient (21,22). We henceforth omit the index t from $\boldsymbol{X}_i$ and Y_i for simplicity of presentation. The parameters of the model are the intercept β_0, the time-fixed coefficients $\beta_1, \ldots, \beta_l$, and the time-varying coefficients $\beta_{l+1}(t), \ldots, \beta_p(t)$. The model's interpretation is that $e^{\beta_j(t)}$ is the adjusted odds ratio (OR) between the resistance and a unit change in the jth variable at time t.

We modelled $\beta_{l+1}(t), \ldots, \beta_p(t)$ by penalized cubic splines with 40 equi-spaced knots, allowing them to be smooth functions of time. Smoothing parameters were estimated via Restricted Maximum Likelihood (REML), and subsequently β_0 and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_l, \beta_{l+1}(t), \ldots, \beta_p(t))^T$ were estimated by minimizing the penalized model deviance (26).

Penalized splines can be expressed from a Bayesian perspective, which allows deriving credible intervals for them (26), while having frequentist coverage probabilities (27–29). Models including different sets of time-varying and time-fixed covariates were compared by a modified version of the Akaike Information Criterion (AIC) that accounts for the uncertainty in the smoothing parameter estimation (30,31). The analysis was performed in R version 3.6.1 using the packages *mgcv* version 1.8-28 (28), *tableone* version 0.10.0 and *mgcViz* version 0.1.6.

The potential dependence between resistance results of bacteria from the same patient was addressed using random intercepts. The incorporation of random intercepts did not change the results substantially, and therefore the results of random-intercept models are presented in the supporting information.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Data-driven simulations

Estimated time-varying covariate-outcome relationships might misrepresent the true underlying causal effect. To further study this, two data-driven simulation studies were conducted. The simulations were designed with a confounding mechanism that affects coefficient estimates while following a hypothetical yet relevant clinical scenario.

In each simulation study, a Data Generating Mechanism (DGM) was designed to set up a scenario illustrating the challenges of interpreting estimated time-varying coefficients. The DGM determined hypothetical relationships between the 20 covariates previously included in our models and the outcome variable of gentamicin resistance. The simulations were data-driven in the sense that the values of the covariates in the DGM were taken from our data. Hence the simulated data structure, including the sample size, were comparable to our data. In each simulation scenario, all but one of the coefficients used to simulate the covariate-outcome relationships were the estimates from the models fitted to the original data. In addition, a hypothetical binary covariate denoted *Community use* was created and included in the DGM of both scenarios, and its values were sampled in each iteration of the simulation. We did not have access to data about antibiotic use outside the hospital. Therefore, we designated *Community use* to hypothetically indicate whether a patient was prescribed antibiotics outside the hospital during the year prior to their susceptibility test. The conditional probability of *Community use* was determined using Bayes' theorem as detailed in the Results section.

Each simulation study consisted of 500 iterations. In each iteration, 11,592 gentamicin resistance results were drawn from the DGM. Then, a time-varying coefficient model with one time-varying coefficient (keeping the rest of the coefficients time-fixed) was fitted using the actual covariates in the data, the simulated values of Community use and resistance results. Finally, the average

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

of the 500 time-varying coefficient estimates alongside their empirical 2.5% and 97.5% quantiles were calculated, as well as the average estimate and empirical 2.5% and 97.5% quantiles from an analogous standard (time-fixed) logistic regression model fitted to the same simulated data.

Results and discussion

Time-varying coefficient models applied to clinical data

Using our clinical data, we modelled the probability of resistance to gentamicin as a function of 20 covariates that were selected based on prior knowledge (32), and are detailed in Tables S1 and S2. We fitted one standard logistic regression model, and four logistic time-varying coefficient models, each with a different single time-varying coefficient.

The four coefficients allowed to vary in time correspond to different patient-level clinical and demographic information: whether the patient's culture was drawn at the Emergency Room (ER) (*Culture drawn at ER*); the patient's sex (*Male*); hospital use of relevant antibiotics during the prior 6 months (*Used antibiotics*), and whether the patient had a Methicillin-resistant *Staphylococcus aureus* (MRSA) infection during the prior year (*Had MRSA)*. Relevant antibiotics were considered those previously shown to have direct cross-resistance links with gentamicin (19). We correspondingly refer to these four models as the *Culture drawn at ER* model, *Male* model, *Used antibiotics* model and *Had MRSA* model. For example, in the *Culture drawn at ER* model, the coefficient of *Culture drawn at ER* was allowed to vary in time, as expressed by having $\beta_1(t)$ (and not β_1) in the following model equation,

$$\begin{aligned} logit[p_t(Y_i = 1|\boldsymbol{X_i})] &= \beta_0 + \beta_1(t) Culture\ drawn\ at\ ER_i + \beta_2 Male_i \\ &+ \beta_3\, Used\ antibiotics_i + \beta_4\, Had\ MRSA_i + \widetilde{\boldsymbol{\beta}}^T \widetilde{\boldsymbol{X}}_i, \end{aligned} \tag{1}$$

where i stands for the ith observation, Y_i is the resistance, $\boldsymbol{X_i}$ is the vector of all covariates, β_0 is the intercept, $\beta_1(t), \beta_2, \beta_3, \beta_4$ are the coefficients of the listed covariates, and $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{X}}_i$ are the

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

vectors of the rest of the 16 coefficients and covariates, respectively.

Table 2 and Figure 1 present estimated coefficients of selected covariates from the four time-varying models and from a standard logistic model using the same 20 covariates. The AIC values (Table 2) suggest that allowing the coefficients of *Culture drawn at ER* and *Male* to vary with time is preferable to keeping them constant. The estimates of these coefficients vary substantially with time (Figure 1), suggesting that perhaps both should be allowed to simultaneously vary with time. Indeed, a model that allowed the coefficients of both *Culture drawn at ER* and *Male* to vary with time had an AIC score of 8970.2, a score lower than the AIC scores of the other five models we fitted (Table 2). The estimates from this model were similar to the estimates from the *Culture drawn at ER* and *Male* models (data not shown). The estimated coefficients which were kept time-fixed in all models were consistent with prior evidence (32) (Tables S3-S7).

	Standard logistic model	***Culture drawn at ER* model**	***Male* model**	***Used antibiotics* model**	***Had MRSA* model**
Covariate	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**
Culture drawn at ER	1.41 (0.91-2.19)	TV	1.38 (0.88-2.13)	1.41 (0.91-2.19)	1.41 (0.91-2.19)
Male	1.32 (1.18-1.48)	1.32 (1.18-1.47)	TV	1.32 (1.18-1.48)	1.32 (1.18-1.48)
Used antibiotics	1.26 (1.11-1.43)	1.26 (1.11-1.44)	1.25 (1.1-1.42)	TV	1.26 (1.11-1.43)
Had MRSA	1.49 (1.23-1.79)	1.49 (1.23-1.80)	1.51 (1.25-1.82)	1.49 (1.23-1.80)	TV
Times hospitalized past year (/3)	1.24 (1.11-1.38)	1.24 (1.11-1.38)	1.24 (1.11-1.39)	1.25 (1.12-1.39)	1.24 (1.11-1.38)
Nosocomial	1.26 (1.12-1.43)	1.27 (1.12-1.44)	1.27 (1.12-1.44)	1.26 (1.11-1.43)	1.26 (1.11-1.43)
Arrived from an institution	1.43 (1.23-1.66)	1.44 (1.24-1.67)	1.43 (1.24-1.67)	1.44 (1.24-1.67)	1.43 (1.23-1.66)

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(AIC)	(8990.1)	(8979.9)	(8981.0)	(8991.3)	(8991.3)
ΔAIC	0	-10.2	-9.1	+1.2	+1.2

Table 2. Selected coefficient estimates and AIC values for the five fitted models. In the standard logistic model, all coefficients were time-fixed. Each time-varying coefficient model was named after its time-varying coefficient, e.g. in the *Culture drawn at ER* model only the coefficient of *Culture drawn at ER* was time-varying. OR: odds ratio; CI: confidence interval; TV: the coefficient was time-varying; (/3): results presented per 3-unit increase.

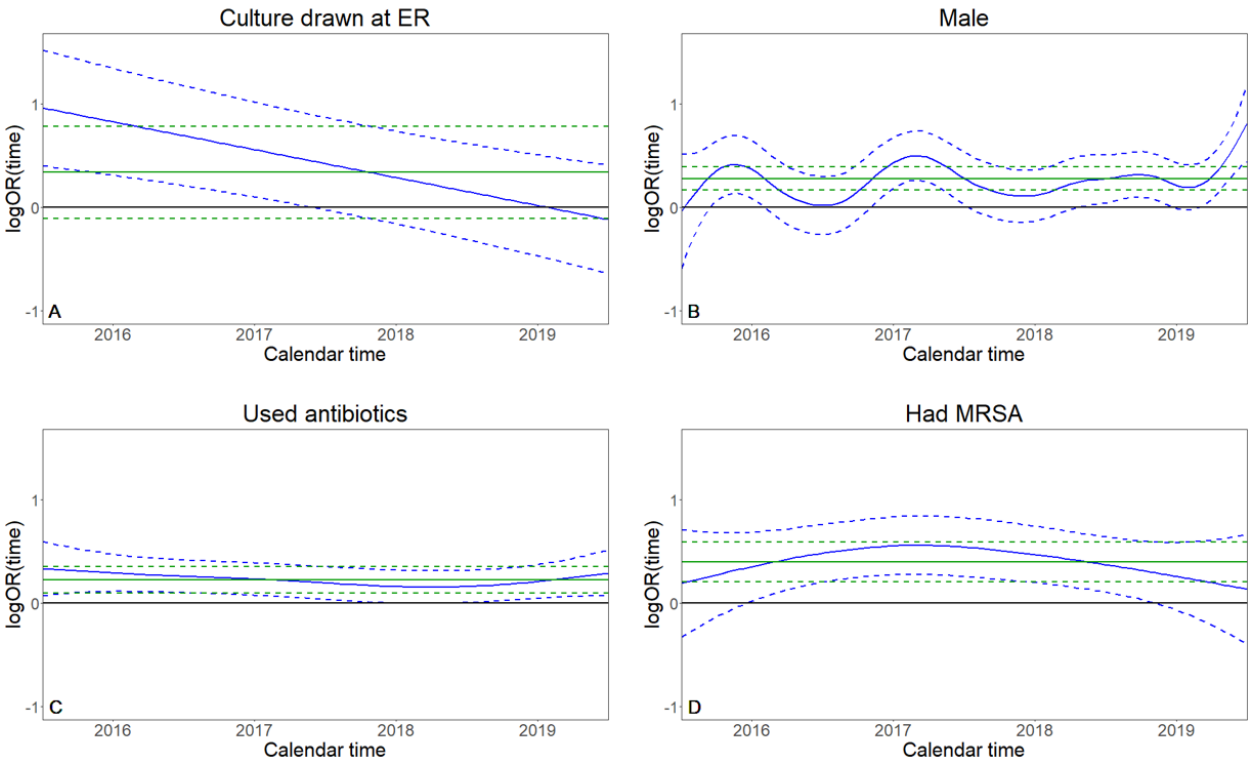

Figure 1. Estimates of the time-fixed (green) and time-varying (blue) coefficients from the single time-fixed and the four time-varying coefficient models. Correspondingly, 95% Bayesian credible intervals and standard 95% confidence intervals are given in dashed lines. The horizontal axis represents the time in days between 2016 and 2019.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Considerable differences were observed between some of the estimated coefficients in the time-varying models and the standard logistic model (Figure 1A-B). While the standard logistic model yielded a positive coefficient estimate for *Culture drawn at ER*, the *Culture drawn at ER* model estimated it as positive only during 2016-18, with zero outside the credible interval only during 2016 to mid-2017, and with a decreasing linear trend towards the null/negative values with time (Figure 1A). Note that the obtained linear shape in Figure 1A was not a-priori imposed but resulted from the REML fitting procedure. The estimated *Male* coefficient also fluctuated with time in the *Male* model, but unlike the coefficient of *Culture drawn at ER,* no clear trend was observed (Figure 1B). When a random intercept was included in the *Male* model, a clearer mildly increasing trend was observed (Figure S1B). Of note is that both the time-varying and the standard logistic model estimated the coefficient of *Male* as positive during 2016-19.

On the other hand, the estimates of the time-varying coefficients in the *Used antibiotics* and *Had MRSA* models were similar to their corresponding estimates in the standard time-fixed logistic model, both in terms of point estimates and uncertainty bounds (confidence intervals for the time-fixed logistic model and credible intervals for the logistic time-varying coefficient models). With that being said, the estimated time-varying coefficient of *Had MRSA* slightly increased during 2016-2017 and then slightly decreased until the end of the study period.

The above-described results align with the arguments we pose in this work. Both *Used antibiotics* and *Had MRSA* are biologically plausible causes for antibiotic resistance. Prior antibiotic use causes resistance via evolutionary selective pressure (33), while a previous MRSA infection can imply remnants of resistant bacteria (34,35). Thus, the estimated coefficients of *Used antibiotics* and *Had MRSA* demonstrate how time-stable relationships may align with causal interpretation. Conversely, it is biologically unlikely that *Culture drawn at ER* and *Male* are direct causes of

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

resistant infections. Therefore, the estimated time-varying relationships of these covariates with resistance are more plausibly explained by confounding.

We now turn to describe and illustrate such confounding mechanisms via data-driven simulations.

Exploring drivers of time-varying coefficient estimates via simulations

First simulation study: Recreating a biased time-varying coefficient estimate

In the first simulation study, we recreated the time-varying coefficient estimate of *Culture drawn at ER* from the *Culture drawn at ER* model (Figure 1A), even though data were simulated such that *Culture drawn at ER* had no causal effect on resistance. That is, the underlying relationship was null. In this hypothetical scenario, the relationship between *Culture drawn at ER* and resistance was confounded by *Community use*, in a time-varying manner. In our scenario, we can hypothesize that increased awareness of antibiotic resistance led doctors treating outside the hospital to prescribe a decreasing number of antibiotics in the community during 2016-19, as was observed in various settings (36,37). As a result, the association between *Culture drawn at ER* and *Community use* over time changed, so that *Culture drawn at ER* was a good surrogate for *Community use* at the beginning of 2016, but deteriorated over time.

The DGM for this scenario employed a standard time-fixed logistic model for the probability of resistance. The DGM included the same 20 covariates as the five models described in the previous subsection, apart from *Culture drawn at ER* which was replaced by *Community use*. For all covariates other than the artificially created *Community use*, the coefficients in the DGM were the estimates from the *Culture drawn at ER* model. We set the coefficient of *Community use* to two, as antibiotic use is a known cause for antibiotic resistance (33). In each simulation iteration, we simulated the resistance results according to a standard time-fixed logistic model

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

$$logit[p_t(Y_i = 1|\boldsymbol{X_i})] = \beta_0 + 2Community\ use_i + \beta_2 Male_i + \beta_3\ Used\ antibiotics_i + \beta_4\ Had\ MRSA_i + \widetilde{\boldsymbol{\beta}}^T\widetilde{\boldsymbol{X}}_i. \quad (2)$$

Using Bayes' theorem, we set the probability of *Community use* as time-varying,

$$p_t(Community\ use_i = 1|Culture\ drawn\ at\ ER_i = k) = \frac{p(Culture\ drawn\ at\ ER_i = k|Community\ use_i = 1)p_t(Community\ use_i = 1)}{p_t(Culture\ drawn\ at\ ER_i = k)}, \quad (3)$$

for $k = 0,1$. For simplicity, we chose a linear decrease of antibiotic prescription in the community over time, $p_t(Community\ use_i = 1) = max\left\{0.1, 0.4 - \frac{t}{4258.333}\right\}$, where the slope and intercept were chosen such that from $t = 1279$ (3.5 years) we obtained $p_t(Community\ use_i = 1) = 0.1$. The resulting approximate density of $p_t(Community\ use_i = 1)$ in the data is shown in Figure S2. In our dataset, approximately 10% of the patients' cultures were sampled at the ER. In this hypothetical scenario, patients who were prescribed antibiotics in the community were on average "less healthy", and thus more likely to have arrived at the ER and had their bacterial culture drawn there, than patients who did not receive antibiotics in the community. Hence, we set $p(Culture\ drawn\ at\ ER_i = 1|Community\ use_i = 1) = 0.2$. Finally, we estimated $p_t(Culture\ drawn\ at\ ER_i = 1)$ from the original data using a time-varying intercept-only model (Figure S3).

In each simulation iteration, we estimated a logistic time-varying coefficient model with *Culture drawn at ER* being the only time-varying coefficient (Methods). The average estimate of the time-varying coefficient of *Culture drawn at ER* across simulations (Figure 2) emulated the estimated coefficient obtained from the *Culture drawn at ER* model fitted to the original data (Figure 1A). To explain this, note that at the beginning of 2016 *Culture drawn at ER* was a good surrogate for *Community use*, and the average estimated time-varying coefficient was close to the actual coefficient of *Community use* in Equation (2) (Figure 2). Then, as the association between the

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

two covariates diminished over time due to the trend of decrease in *Community use*, the average time-varying coefficient estimate of *Culture drawn at ER* diverged from the coefficient of *Community use* towards zero, its true value under the DGM (Equation (2)). Thus, this simulation study demonstrated how a confounding mechanism that varies with time can result in a misleading estimate.

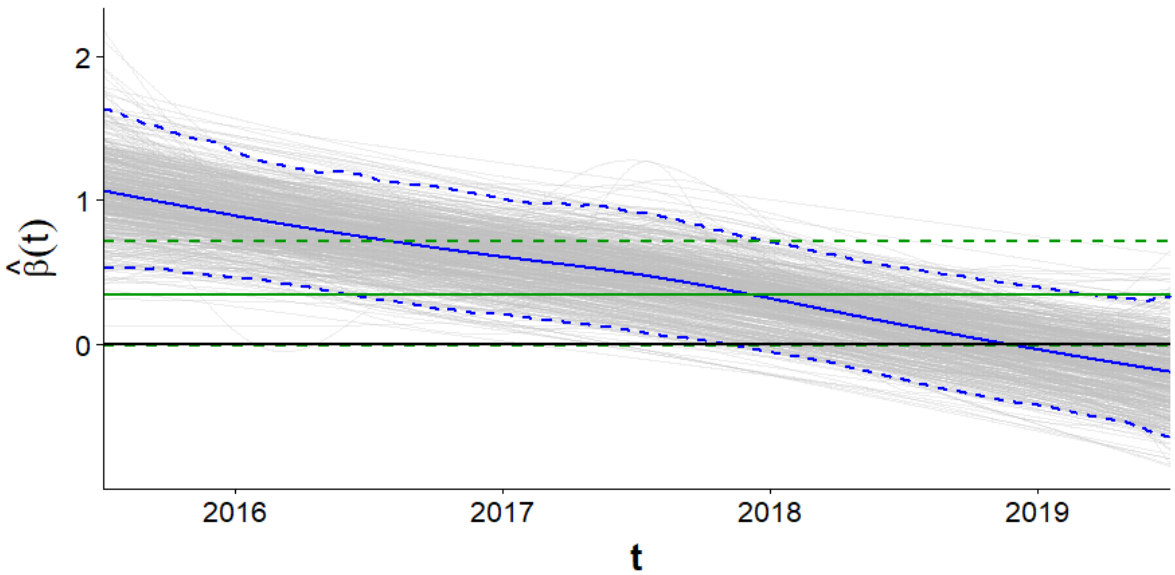

Figure 2. Average estimates of the time-fixed (green) and time-varying (blue) coefficients from the 500 time-fixed and time-varying coefficient models from the first simulation study. Correspondingly, 2.5% and 97.5% quantiles of coefficients estimates are in dashed green and blue. The 500 time-varying coefficient estimates are depicted in grey. The horizontal axis represents the time in days between 2016 and 2019.

For comparison, the average coefficient estimate of *Culture drawn at ER* from the standard time-fixed logistic model (Figure 2) was also biased. Its average was 0.35 (2.5%, 97.5% quantiles: -0.01, 0.72), while in practice *Culture drawn at ER* had no effect on the resistance. Hence time-fixed models are also subject to bias in such a scenario.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Second simulation study: Recreating a biased time-stable coefficient estimate

In the second data-driven simulation, we considered a causal relationship that varied considerably with time but was approximately estimated as time-stable, even though the estimated coefficient was allowed to vary in time. To obtain this approximate time-stable estimate, the confounding mechanism had to vary with time in a way that near-perfectly negated the underlying time-varying relationship. Hence, it demanded parameter values that are unlikely to be realistic in our context.

In this simulation scenario, the relationship between *Had MRSA* and resistance was confounded by *Community use* in a time-varying manner. The time-varying effect of *Had MRSA* on resistance was set as positive, and decreased over time. The decrease over time can be motivated through changes in the bacterial population of *Staphylococcus aureus*, as it can affect the resistance of other bacteria via horizontal gene transfer (HGT) (34). As in the first simulation, we mimicked increased awareness of antibiotic resistance that led community doctors to decrease antibiotic prescription over time. As a result, the association between *Had MRSA* and *Community use* changed over time, so that *Had MRSA* started as a good surrogate for *Community use* but deteriorated over time. We designed the change in the association between *Had MRSA* and *Community use* to specifically induce bias into the estimate of the time-varying coefficient of *Had MRSA*, such that a constant estimate was obtained.

We set the coefficient of *Community use* to 1.25, and the time-varying coefficient of *Had MRSA* to $\beta_{Had\ MRSA}(t) = min\left\{0.5, \frac{t}{2000}\right\}$. Then, in each iteration we simulated the resistance results according to the following DGM,

$$logit[p_t(Y_i = 1|\boldsymbol{X}_i)] = \beta_0 + 1.25 Community\ use_i + \beta_2 Culture\ drawn\ at\ ER_i$$
$$+\beta_3 Male_i + \beta_4\ Used\ antibiotics_i + \beta_{Had\ MRSA}(t) Had\ MRSA_i + \widetilde{\boldsymbol{\beta}}_i^T \widetilde{\boldsymbol{X}}_i.$$

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Again, we set the probability of *Community use* using Bayes' Theorem,

$$p_t(Community\ use_i = 1 | Had\ MRSA_i = k) = \frac{p(Had\ MRSA_i = k | Community\ use_i = 1) p_t(Community\ use_i = 1)}{p(Had\ MRSA_i = k)}, \quad (4)$$

for $k = 0,1$. As in the previous simulation study, the values of each component in Equation (4) followed a clinical scenario in which antibiotic community use linearly decreased during 2016-19. Therefore, we set $p_t(Community\ use_i = 1) = max\left\{0, 0.4 - \frac{t}{2500}\right\}$. In our dataset, approximately 6% of the patients have previously had an MRSA infection, hence we set $p(Had\ MRSA_i = 1) = 0.06$. Finally, we assumed that patients who were prescribed antibiotics in the community are on average more likely to have had an MRSA infection than those who were not, and set $p(Had\ MRSA_i = 1 | Community\ use_i = 1) = 0.12$.

In each iteration, we estimated a logistic time-varying coefficient model with the 20 available covariates in which the coefficient of *Had MRSA* was the only time-varying coefficient (Methods section). The obtained average coefficient estimate (Figure 3) was time-stable, and resembled the coefficient estimates from the *Used antibiotics* and *Had MRSA* models in the original data (Figure 1C-D).

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

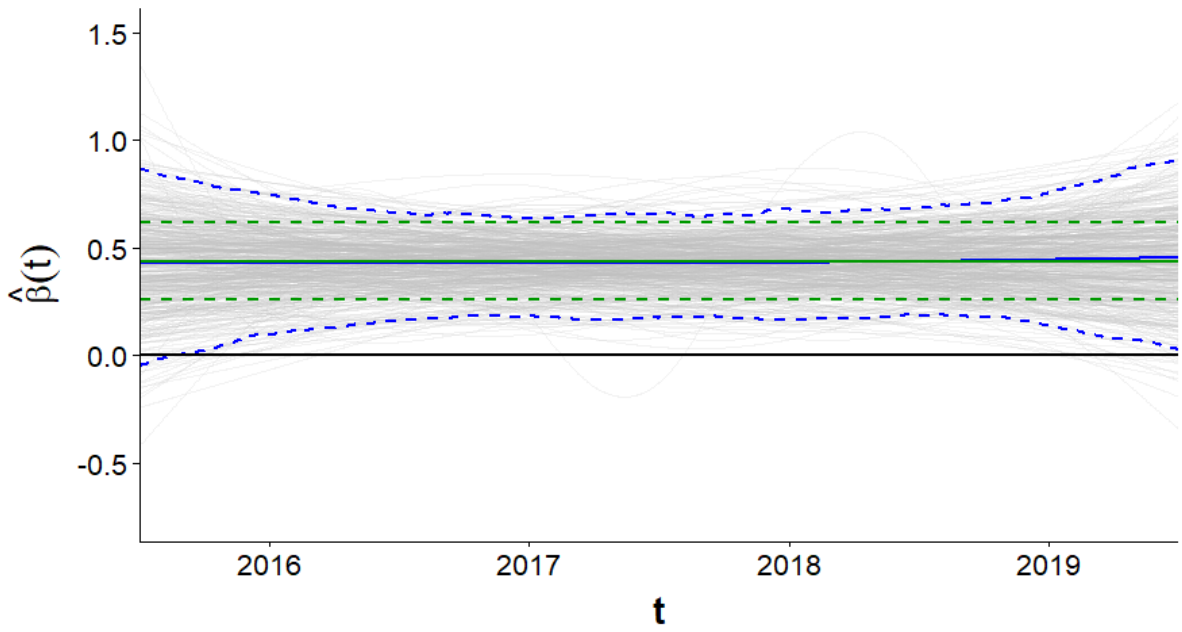

Figure 3. Average estimates of the time-fixed (green) and time-varying (blue) coefficients from the 500 time-fixed and time-varying coefficient models. Correspondingly, 2.5% and 97.5% quantiles of coefficients estimates are in dashed green and blue. The 500 time-varying coefficient estimates are depicted in grey. The horizontal axis represents the time in days between 2016 and 2019.

As in the first simulation study, in each iteration we also fitted a standard time-fixed logistic model. The average estimate of the time-fixed coefficient of *Had MRSA* was also biased (Figure 3).

Conclusions

Many epidemiological settings, and especially those involving infectious diseases, are characterized by time-varying relationships. However, epidemiological studies often analyze data as static, essentially averaging observed associations across time and neglecting to account for the dynamical nature of the studied system. Using a combination of an antibiotic resistance

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

related clinical dataset and data-driven simulations, we have demonstrated that time-varying coefficient models can reveal phenomena otherwise obscured by time-fixed models.

We showcased how the use of time-varying coefficient models may help unravel causal covariate-outcome relationships. As we have shown, estimates that are time-stable can coincide with prior knowledge suggesting that the underlying relationships are causal. Conversely, time-varying estimates may direct researchers to inquire the underlying causal structure leading to the obtained time-varying relationships.

Time-varying coefficient estimates can result from a mixture of underlying covariate-outcome relationships and confounding mechanisms that are time-fixed and/or time-varying. It is possible the underlying covariate-outcome relationships are time-fixed (or non-existent), while time-varying confounding leads to a time-varying estimate; we demonstrated such a scenario in our first data-driven simulation study. On the other hand, time-stable coefficient estimates are more likely the consequence of underlying time-fixed (or non-existent) covariate-outcome relationships, that are possibly confounded in a time-fixed manner. While it is technically possible that time-varying confounding will lead to a time-stable estimate of an underlying time-varying covariate-outcome relationship, this scenario is unlikely. Such a scenario requires the variations in the underlying relationship and in the confounding mechanism to approximately negate each other, as we illustrated in the second data-driven simulation study. This is implausible and hence can often be ruled out, though it is advised to do so based on subject-matter expertise.

From a clinical standpoint, knowledge about causal relationships can help researchers understand the biological mechanisms behind them, and possibly result in new interventions. This knowledge can also be exploited to improve existing predictive models for antibiotic resistance (e.g. (38–41)). By emphasizing covariates with causal effects on resistance, predictions from such

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

models might prove more stable and robust to generalizations over different time periods or locations.

In the context of antibiotic resistance, future research should extend efforts of further identifying causal relationships between risk factors and antibiotic resistance using time-varying models. Development of a theoretical framework for assessing causal effects in the field of antibiotic resistance (i.e., based on potential outcomes and/or directed acyclic graphs), while accounting for variations with time, would be clinically and methodologically valuable.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Funding

This work was supported by a grant from the Tel Aviv University Center for AI and Data Science (TAD).

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Centers for Disease Control and Prevention (U.S.). Antibiotic resistance threats in the United States, 2019 [Internet]. Centers for Disease Control and Prevention (U.S.); 2019 Nov [cited 2021 Oct 25]. Available from: https://stacks.cdc.gov/view/cdc/82532

2. Organization WH, others. Global antimicrobial resistance and use surveillance system (GLASS) report: 2021.

3. Bottery MJ, Pitchford JW, Friman V-P. Ecology and evolution of antimicrobial resistance in bacterial communities. ISME J. 2021 Apr;15(4):939–48.

4. Hughes D, Andersson DI. Evolutionary Trajectories to Antibiotic Resistance. Annu Rev Microbiol. 2017 Sep 8;71(1):579–96.

5. Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. Pharmacy and therapeutics. 2015 Apr;40(4):277.

6. Vernaz N, Huttner B, Muscionico D, Salomon J-L, Bonnabry P, Lopez-Lozano JM, et al. Modelling the impact of antibiotic use on antibiotic-resistant Escherichia coli using population-based data from a large hospital and its surrounding community. J Antimicrob Chemother. 2011 Apr 1;66(4):928–35.

7. Bell BG, Schellevis F, Stobberingh E, Goossens H, Pringle M. A systematic review and meta-analysis of the effects of antibiotic consumption on antibiotic resistance. BMC Infect Dis. 2014 Dec;14(1):13.

8. Holmes AH, Moore LSP, Sundsfjord A, Steinbakk M, Regmi S, Karkey A, et al. Understanding the mechanisms and drivers of antimicrobial resistance. The Lancet. 2016 Jan;387(10014):176–87.

9. European Centre for Disease Prevention and Control (ECDC), European Food Safety Authority (EFSA), European Medicines Agency (EMA). ECDC/EFSA/EMA second joint report on the integrated analysis of the consumption of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from humans and food-producing animals. EFSA J [Internet]. 2017 Jul [cited 2021 Nov 23];15(7). Available from: https://data.europa.eu/doi/10.2903/j.efsa.2017.4872

10. Canizalez-Roman A, Gonzalez-Nuñez E, Vidal JE, Flores-Villaseñor H, León-Sicairos N. Prevalence and antibiotic resistance profiles of diarrheagenic Escherichia coli strains isolated from food items in northwestern Mexico. Int J Food Microbiol. 2013 Jun;164(1):36–45.

11. European Food Safety Authority, European Centre for Disease Prevention and Control. The European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2016. EFSA J [Internet]. 2018 Feb [cited 2021 Nov 23];16(2). Available from: https://data.europa.eu/doi/10.2903/j.efsa.2018.5182

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

12. Winglee K, Howard AG, Sha W, Gharaibeh RZ, Liu J, Jin D, et al. Recent urbanization in China is correlated with a Westernized microbiome encoding increased virulence and antibiotic resistance genes. Microbiome. 2017 Dec;5(1):121.

13. Peng F, Guo Y, Isabwe A, Chen H, Wang Y, Zhang Y, et al. Urbanization drives riverine bacterial antibiotic resistome more than taxonomic community at watershed scale. Environ Int. 2020 Apr;137:105524.

14. Obolski U, Lourenço J, Thompson C, Thompson R, Gori A, Gupta S. Vaccination can drive an increase in frequencies of antibiotic resistance among nonvaccine serotypes of *Streptococcus pneumoniae*. Proc Natl Acad Sci. 2018 Mar 20;115(12):3102–7.

15. the WHO Guidelines Development Group, Storr J, Twyman A, Zingg W, Damani N, Kilpatrick C, et al. Core components for effective infection prevention and control programmes: new WHO evidence-based recommendations. Antimicrob Resist Infect Control. 2017 Dec;6(1):6.

16. Review on Antimicrobial Resistance. Infection prevention, control and surveillance: limiting the development and spread of drug resistance. Review on Antimicrobial Resistance; 2016.

17. Carmeli Y, Castro J, Eliopoulos GM, Samore MH. Clinical Isolation and Resistance Patterns of and Superinfection with 10 Nosocomial Pathogens after Treatment with Ceftriaxone versus Ampicillin-Sulbactam. Antimicrob Agents Chemother. 2001 Jan;45(1):275–9.

18. Kaye KS, Kanafani ZA, Dodds AE, Engemann JJ, Weber SG, Carmeli Y. Differential Effects of Levofloxacin and Ciprofloxacin on the Risk for Isolation of Quinolone-Resistant *Pseudomonas aeruginosa*. Antimicrob Agents Chemother. 2006 Jun;50(6):2192–6.

19. Cherny SS, Nevo D, Baraz A, Baruch S, Lewin-Epstein O, Stein GY, et al. Revealing antibiotic cross-resistance patterns in hospitalized patients through Bayesian network modelling. :10.

20. Chowers M, Zehavi T, Gottesman B-S, Baraz A, Nevo D, Obolski U. Estimating the cost of antibiotic use on future collateral resistance: a retrospective comparison of cefuroxime versus cefazolin and amoxicillin/clavulanate [Internet]. Infectious Diseases (except HIV/AIDS); 2022 Jan [cited 2022 Jan 24]. Available from: http://medrxiv.org/lookup/doi/10.1101/2022.01.10.22269003

21. Hastie T, Tibshirani R. Varying-Coefficient Models. J R Stat Soc Ser B Methodol. 1993 Sep;55(4):757–79.

22. Hastie TJ, Tibshirani RJ. Generalized additive models. Routledge; 2017 Oct 19.

23. Berger P, Pascal L, Sartor C, Delorme J, Monge P, Ragon CP, et al. Generalized Additive Model Demonstrates Fluoroquinolone Use/Resistance Relationships for Staphylococcus aureus. Eur J Epidemiol. 2004;19(5):453–60.

24. Kiffer CR, Camargo EC, Shimakura SE, Ribeiro PJ, Bailey TC, Pignatari AC, et al. A spatial approach for the epidemiology of antibiotic use and resistance in community-based studies: the emergence of urban clusters of Escherichia coli quinolone resistance in Sao Paulo, Brasil. Int J Health Geogr. 2011 Dec;10(1):17.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

25. Sommerstein R, Atkinson A, Lo Priore EF, Kronenberg A, Marschall J, Burnens A, et al. Characterizing non-linear effects of hospitalisation duration on antimicrobial resistance in respiratory isolates: an analysis of a prospective nationwide surveillance system. Clin Microbiol Infect. 2018 Jan;24(1):45–52.

26. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models: Estimation of Semiparametric Generalized Linear Models. J R Stat Soc Ser B Stat Methodol. 2011 Jan;73(1):3–36.

27. Wahba G. Bayesian “Confidence Intervals” for the Cross-Validated Smoothing Spline. J R Stat Soc Ser B Methodol. 1983 Sep;45(1):133–50.

28. Wood SN. Generalized Additive Models: An Introduction with R SECOND EDITION. :497.

29. Marra G, Wood SN. Coverage Properties of Confidence Intervals for Generalized Additive Model Components: Coverage properties of GAM intervals. Scand J Stat. 2012 Mar;39(1):53–74.

30. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. 1974 Dec;19(6):716–23.

31. Wood SN, Pya N, Säfken B. Smoothing Parameter and Model Selection for General Smooth Models. J Am Stat Assoc. 2016 Oct 1;111(516):1548–63.

32. Chatterjee A, Modarai M, Naylor NR, Boyd SE, Atun R, Barlow J, et al. Quantifying drivers of antibiotic resistance in humans: a systematic review. Lancet Infect Dis. 2018 Dec;18(12):e368–78.

33. Davies J, Davies D. Origins and Evolution of Antibiotic Resistance. Microbiol Mol Biol Rev. 2010 Sep;74(3):417–33.

34. McCarthy AJ, Lindsay JA. The distribution of plasmids that carry virulence and resistance genes in Staphylococcus aureus is lineage associated. BMC Microbiol. 2012;12(1):104.

35. Mattner F, Biertz F, Ziesing S, Gastmeier P, Chaberny IF. Long-term persistence of MRSA in re-admitted patients. Infection. 2010 Oct;38(5):363–71.

36. King LM, Bartoces M, Fleming-Dutra KE, Roberts RM, Hicks LA. Changes in US Outpatient Antibiotic Prescriptions From 2011–2016. Clin Infect Dis. 2019 Mar 16;ciz225.

37. Holstiege J, Schulz M, Akmatov MK, Kern WV, Steffen A, Bätzing J. The Decline in Outpatient Antibiotic Use. Dtsch Ärztebl Int. :15.

38. Oonsivilai M, Mo Y, Luangasanatip N, Lubell Y, Miliya T, Tan P, et al. Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children’s hospital in Cambodia. Wellcome Open Res. 2018 Oct 10;3:131.

39. Lewin-Epstein O, Baruch S, Hadany L, Stein GY, Obolski U. Predicting Antibiotic Resistance in Hospitalized Patients by Applying Machine Learning to Electronic Medical Records. Clin Infect Dis. 2021 Jun 1;72(11):e848–55.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270156; this version posted January 31, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

40. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med. 2019 Apr 4;380(14):1347–58.

41. Yelin I, Snitser O, Novich G, Katz R, Tal O, Parizade M, et al. Personal clinical history predicts antibiotic resistance of urinary tract infections. Nat Med. 2019 Jul;25(7):1143–52.