

Personal clinical history predicts antibiotic resistance in urinary tract infections

Idan Yelin, Olga Snitser, Gal Novich, Rachel Katz, Ofir Tal, Miriam Parizade, Gabriel Chodick, Gideon Koren, Varda Shalev, Roy Kishony*

Abstract

The prevalence of antibiotic resistance in urinary tract infections (UTIs) often renders the prescribed antimicrobial treatment ineffective, highlighting the need for personalized prediction of resistance at time of care. Here, crossing a 10-year longitudinal dataset of over 700,000 community-acquired UTIs with over 6,000,000 personally-linked records of antibiotic purchases, we show that the resistance profile of infections can be predicted based on patient-specific demographics and clinical history. Age, gender, and retirement home residence had strong, yet differential and even non-monotonic, associations with resistance to different antibiotics. Resistance profiles were also associated with the patient's records of past urine samples and antibiotic usage, with these associations persisting for months and even longer than a year. Drug usage selected specifically for its own cognate resistance, which led indirectly, through genetic linkage, also to resistance to other, even mechanistically unrelated, drugs. Applying machine learning models, these association patterns allowed good personalized predictions of resistance, which could inform and better optimize empirical prescription of antibiotics.

The resistance of bacterial pathogens to commonly used antibiotics is a growing public health concern, threatening the efficacy of life-saving antibiotic drugs ^{1,2}. Antibiotic use and misuse can benefit resistant strains, exacerbating the problem over time ³⁻⁵. At the single patient level, the efficacy of antimicrobial treatment is critically dependent on correctly matching antibiotic choice to the specific susceptibilities of the pathogen ⁶⁻⁸. Ideally, correct prescription should be based on direct measurement of the antibiotic susceptibilities of the infecting pathogen. In practice, though, to save time and resources, drugs are often prescribed empirically in absence of culture susceptibility measurements, risking incorrect and ineffective treatment.

This problem is of particular importance in Urinary Tract Infections (UTIs) which are often treated empirically despite substantial levels of resistance. UTIs are among the most common bacterial infections, with over 150 million annual cases globally ⁹. One of three women will have at least one symptomatic UTI by age 24, and more than one-half will be affected during their lifetime ¹⁰. Treatment of these infections accounts for about 8% of non-hospital usage of antibiotics, often as part of empirical prescription ^{11,12}. The common etiological agents of UTIs are diverse, including *Escherichia coli*, *Klebsiella pneumoniae* and *Proteus mirabilis*, as well as gram-positive bacteria such as *Enterococcus faecalis* ^{13–18}. These pathogens are often resistant to several antibiotics, with resistance rates of infections exceeding 20% for commonly used drugs ^{14,17,19}, emphasizing the challenge of choosing the most appropriate antibiotic treatment for each patient ²⁰.

The risk of an infection being resistant to different antibiotics can be correlated with patient demographics and clinical and epidemiological factors. Known demographic factors associated with resistance include older age ²¹, gender ²², ethnicity ^{23–26} and residence in a retirement home ²². Known clinical and epidemiological factors associated with resistance include presence of a urinary catheter ^{18,22,27}, immunodeficiency ²², diabetes ²² and travel to developing countries ²⁵. Notably, most of these associations were identified based on small patient cohorts, typically with high frequencies of antibiotic resistant infections, such as retirement homes, rehabilitation centers, or hospitals.

Beyond the patient's demographics and current clinical factors, antibiotic resistance can also be associated with the patient's past clinical history, including recurrent UTIs and hospitalizations as well as past resistance profiles and antibiotic prescriptions. Risk of resistance to specific drugs have been shown to increase for patients with recurrent UTIs ^{22,26,28} and past hospitalizations ^{22,29}. Focusing on hospital settings, studies have further shown that past resistance profiles can be used to predict resistance in future infections ³⁰. Availability of antibiotic purchase data reveals patterns of antibiotic use and misuse ³¹ and shows that risk of resistance increases with short-term prior use of antibiotics ^{5,21,22,29,32–34}. At the population level, a recent large-scale study revealed correlations between risk of trimethoprim/sulfa resistance and the volume of past purchases of the same drug (cognate) as well as of other drugs of different pharmaceutical classes (noncognate) ¹⁷. Such associations of usage of a given antibiotics with future resistance to other antibiotics can appear indirectly through genetic linkage among resistance mechanisms, but mathematically resolving direct and indirect selection for resistance has been challenging. Negative associations, where drug use is anti-correlated with resistance, have also been observed, but it has been

difficult to discern the direction of causality^{17,35}. Finally, the time extent of these positive and negative associations of resistance with prior antibiotic usage and prior resistant samples is not well resolved, and it is also unclear whether and how these associations vary across resistances to different antibiotics.

Here, we present a systematic big-data analysis of a large population of UTI patients to unravel predictive features of antibiotic resistance. We analyze a patient-level longitudinal dataset of community and retirement-home acquired UTI cultures collected by Maccabi Healthcare Services (MHS), Israel's second largest Health Maintenance Organizations, serving a diverse population of ~2 million patients. We first analyze correlations between demographic factors and antibiotic resistance. Then, comparing resistance data of multiple infections from the same patient, we unravel long-term “memory” of resistance over time. We also combine these culture records with patient-linked records of antibiotic use to quantify the extent and time of direct and indirect impacts of antibiotic use on resistance at the single-patient level. Finally, we develop machine learning models combining these demographic and historical factors for personalized predictions of resistance.

Results

We retrieved data of all positive urine cultures of MHS patients for the ten-year period between 01-July-2007 and 30-June-2017, as well as patient demographics and record of antibiotic purchases for these patients (Methods). Among all ~2 million MHS patients, there were 711,099 recorded positive urine samples from 315,047 patients total. For each positive sample, one or more bacterial species were isolated and characterized (total of 736,793). The dataset included the species identification of these isolates as well as VITEK2 antibiotic resistances reinterpreted in accordance with CLSI guidelines (Sensitive, Intermediate, and Resistant). For each UTI sample, we also defined “sample resistance” as the maximal resistance for each antibiotic across all isolates from the same sample (96.4% of samples were identified as single species and their resistance profile is simply defined as the resistance profile of their single isolates). In our analysis, we focus on the 16 most frequently measured antibiotic resistances (**Table 1** and **Supplementary Fig. 1**). All of MHS's country-wide clinical tests are performed centrally (Methods), allowing reliable comparison across patients and time.

At the population level, resistance frequencies of different pathogens to different antibiotics varied mildly across time. Three species, *E. coli*, *K. pneumoniae* and *P. mirabilis*, account for 85% of isolates (70%, 10%, 5%, respectively; **Fig. 1a**). These pathogens varied in their resistance profiles (**Fig. 1b**). Notably, for almost any of the 16

considered antibiotics, the chance of resistant infection is significant, indicating that antibiotic treatment efficacy could often be undermined. Notable exceptions are amikacin, meropenem and ertapenem, for which resistance frequencies were negligible (<1% of UTIs); yet these are last-line antibiotics of restricted use. These population-level frequencies of resistance were fairly static over time and only mild changes were observed in certain antibiotics and specific species (**Fig. 1c** and **Supplementary Fig. 2**). The diversity of pathogens and resistance patterns underscores the widely accepted notion that antibiotic prescriptions must be tailored based on the resistance profile of the infection, motivating the development of methods to better predict resistance ²⁰.

The effect of demographics on antibiotic resistance

Considering the distribution of demographic factors across the UTI dataset, we find gender-specific age distributions (**Fig. 2a**). UTIs were much more common for females than males (~88% females, consistent with previous literature ^{9,23}) and had qualitatively different age distributions. For females, UTIs appeared primarily in three distinct demographic groups: young children; reproductive-age women, including pregnant women; and the elderly, including retirement-home patients (**Fig. 2a**, top). In contrast, for males, there were only two peaks: an early-life peak of UTIs, dominated by infants, and a late peak of events for the elderly (**Fig. 2a**, bottom). Considering geography, the dataset spanned multiple cities which varied in their distribution of demographic factors (**Supplementary Fig. 3**). In total, this set of community-acquired UTIs showed diversity of demographic factors, allowing analysis of their impacts on resistance.

Age, gender, pregnancy and residence in retirement home had strong, yet differential, association with resistances to the 16 antibiotics. For each antibiotic, we performed multivariate logistic regression for the odds of resistance ($\eta = P_{Resistance} / (P_{Sensitive} + P_{Intermediate})$) as a function of age (quadratically), gender, retirement home residence, pregnancy, city, date of sampling (quadratically) and season of sampling (Methods: Logistic regression “Demographics” model). For all of the 16 antibiotics, risk of resistance strongly increased with age and with retirement-home residence and decreased for females and pregnancy (**Fig. 2b**; and **Supplementary Table 1** for regression coefficients and 95% CI). The effect of age on the odds of resistance differed widely among the 16 measured antibiotics, ranging from odds ratios (OR) of 2 in trimethoprim/sulfa and amoxicillin to more than 16 in antibiotics such as ciprofloxacin and amikacin (**Fig. 2b** and **Supplementary Table 1**). For most antibiotics, females had lower odds of resistance, yet the odds ratios varied substantially among the different antibiotics (from OR=0.94 CI: 0.92-0.96 for trimethoprim/sulfa to OR=0.12 CI: 0.10-0.15 for meropenem). These lower odds of resistance for females were even

lower with pregnancy (as much as OR=0.46; 95% confidence interval, 0.44-0.49 for ciprofloxacin; **Supplementary Table 1**). While the city of residence, when considered alone (Methods: Logistic regression “City, unadjusted”), was correlated with risk of resistance, these associations stemmed from differences in demographics among cities; they were much reduced in the complete regression model (compare adjusted and unadjusted coefficients, **Supplementary Table 1**). The date of sample had some association with resistance to specific antibiotics, most notably ampicillin and ceftriaxone, while season had almost no effect on resistance for any of the drugs (**Fig. 2b**).

The dependence of resistance on age showed complex and antibiotic-specific patterns. To better understand how the demographic factors interact and what is the effect of age as a continuous variable, we calculated, for each of the 16 antibiotics the odds of resistance η of the urine samples across ten age groups for five categories: men not residing in retirement homes, men residing in retirement homes, women not pregnant and not residing in retirement homes, women in retirement homes, and pregnant women (**Fig. 2c** and **Supplementary Fig. 4**). For some antibiotics, the odds of an infection being resistant were non-monotonic with age, with an additional peak of higher resistance risk at infancy or young age (e.g., nitrofurantoin, piperacillin/tazobactam; **Fig. 2c** and **Supplementary Fig. 4**). The effect of age was not always independent of other variables. We identified interaction between gender and age leading to heterogeneous patterns for males and females (e.g. amikacin, meropenem) and even to opposite effects of gender on odds ratios in specific ages groups (e.g. ciprofloxacin; **Fig. 2c** and **Supplementary Fig. 4**). While, across all antibiotics, resistance was higher for residents of retirement homes, the correlation with age within this group was reversed: the odds of resistance for retirement home residence decreased with age (**Fig. 2c** and **Supplementary Fig. 4**). We concluded that among the different demographic factors associated with risk of resistance, age, gender and residence in retirement homes are the strongest, with resistances to different antibiotics differentially affected by these factors and by interactions among them.

Correlation of resistance among same-patient urine samples

Moving from demographics to clinical history, we analyzed correlations of resistance across same-patient infections, revealing short- and long-term “memory” and a constant patient-specific tendency for resistance. For each antibiotic A, we analyzed all same-patient pairs of samples and calculated the odds ratio for resistance of the second sample given the resistance of the first sample ($OR_{pairs} = (N_{R \rightarrow R}/N_{R \rightarrow S}) / (N_{S \rightarrow R}/N_{S \rightarrow S})$), where the N ’s are number of same-patient sample pairs with the specified resistance

phenotypes; for example, $N_{R \rightarrow S}$ is the number of sample pairs in which the first sample is resistant to antibiotic A and the second sensitive; Methods). Calculating OR_{pairs} as a function of the time difference $t = t_1 - t_2$ between the two samples in each pair, we find that, for all antibiotics, these odds ratios are highest for short time differences between the samples, then decay with two distinct time scales and finally converge, at the longer time differences, to a non-zero asymptotic effect (the odds ratios are well fitted by the sum of a fast exponent, a slow exponent and a constant, $OR_{pairs} \approx C_0 e^{t/\tau_0} + C_m e^{t/\tau_m} + C_\infty$; **Fig. 3a,b** and **Supplementary Fig. 5**). The shortest time scale τ_0 , which is about a week, stems from successive samples likely taken from the same infection. For all antibiotics, resistance of a prior sample within this short time increased the odds of resistance of the second sample by not less than two orders of magnitude (**Fig 3a,b,d**, red). The second time scale τ_m represents much longer “memory” of correlations among samples, lasting more than six months for most antibiotics and even longer than a year for some (**Fig. 3c**). These long-term associations among sample measurements are consistent with repeated infection with the same or similar strains. The effect of this memory component was much more consequential and long lasting than previously thought. The maximal odds ratios of this memory term (C_m) ranged between 6 and 14 for different antibiotics and typically remained larger than 2 even for samples taken a year apart (**Fig 3a,b** and **Supplementary Fig. 5**). At much longer times, this memory decayed, and OR_{pairs} converged to a constant, but interestingly it did not diminish, but rather converged to values larger than 1 (**Fig. 3a,b,d**, green), representing timeless patient specific tendencies for resistance. In sum, we found an antibiotic-specific long-term memory of resistance between infections that can last for months to a year, as well as timeless correlations among even more remote samples, which can greatly improve predictions of resistance.

Association of resistance with past antibiotic purchase

Next, we crossed the infection dataset with patient-linked antibiotic purchase data. For each patient with recorded UTI samples, we retrieved all records of antibiotic purchase made during the twenty year period from 1-Jan-1998 to 30-Jun-2017. For analysis, we used 34 drug groups (Methods; **Supplementary Table 2**). For each UTI sample, we counted the number of purchases made by the same patient of each of the 34 drugs at distinct time intervals prior to the sample (Methods). Then, we applied multivariate logistic regression to correlate resistance to each of the 16 antibiotics with these drug purchase counts (Methods: Logistic regression “Purchase history”; **Fig. 4a** and **Supplementary Fig. 6a**).

We identified strong long-term patient-level associations of resistance with past purchase of both cognate and noncognate antibiotics. For 22 out of the 34 drugs, we identified strong associations of prior drug purchase with future resistance, often lasting for months and even more than a year (**Fig. 4a** and **Supplementary Fig. 6a**). For example, the associations between purchase of ciprofloxacin and its cognate resistance having odds ratios of 1.5 after half a year and remaining as large as 1.2 even two years past purchase (**Fig. 4a**). Some weak negative associations were also identified (ciprofloxacin resistance was negatively correlated with past use of amoxicillin, **Fig. 4a**). Yet, the magnitude of these negative correlations decreased after adjusting for demographics, suggesting that they stemmed indirectly from correlations of purchases and resistance with demographics (Methods: Logistic regression, “Purchase history adjusted for demographics”; **Supplementary Fig. 6b**). Notably, drug purchases were associated not only with their expected cognate resistances. Indeed, use of both first-line antibiotics (such as ciprofloxacin, ofloxacin and levofloxacin) and last-line antibiotics (such as amikacin), increased the risk of a future resistance to a wide range of mechanistically diverse antibiotics. These abundant long-term positive associations between resistances and noncognate drugs did not stem from correlations of purchases and resistance with patient demographics; they remained strong even when adjusting for demographics (**Supplementary Fig. 6b**). Together, these results support strong and long-lasting associations of antibiotic resistance with past use of both cognate and noncognate antibiotics .

Exposing direct drug-to-resistance associations by disentangling linkage among resistances, showed that drug usage specifically selects for its cognate resistance at the single-patient level. Across the sample dataset, resistances to different antibiotics within class and even resistances to antibiotics of different classes were highly correlated (cross resistance; **Supplementary Fig. 7**). These inherent correlations among resistances suggest that observed associations between resistance to a given drug A and past purchase of a different drug B may arise indirectly through selection for resistance B and association between resistance to B and resistance to A. Mathematically discerning these direct from indirect effects is not possible when only a single resistance is considered ^{17,36}. Yet, as our dataset contained measurements of multiple resistances for each sample, we were able to disentangle direct from indirect associations by adjusting the logistic regression for other measured resistances (Methods: Logistic regression “Purchase history adjusted for cross-resistance”). In this cross-resistance adjusted analysis of purchase-resistance associations, the noncognate associations between drug purchases and resistance much diminished and even disappeared while the associations between cognate drug-to-resistance pairs persisted (**Fig. 4b** and **Supplementary Fig. 6c**). For example, considering association between

purchases of trimethoprim/sulfa and ciprofloxacin to their cognate resistances, we observed that the total (unadjusted) effect of cognate drugs remained when adjusting for cross-resistance (**Fig 4c,d**, thick solid vs. thick dashed lines), while the total effect of drugs on their noncognate resistance was much diminished when removing indirect effect (**Fig 4c,d**, thin solid vs. thin dashed lines). Our analysis therefore suggests both direct and indirect selection for resistance at the single-patient level lasting months and even a year following drug use.

Predicting antibiotic resistance at the single-patient single-infection level

Finally, as resistance is strongly associated with demographics, sample history and purchase history, we wanted to determine the predictive power of these factors individually and when combined together. Models of Logistic Regression and Gradient Boosting Decision Trees (GBDT) were trained and tested on temporally separate periods: training period of 9 years from 1-July-2007 to 30-June-2016 and testing period of the following year, from 1-July-2016 to 30-June-2017 (for some antibiotics, training and testing periods were modified to avoid time periods during which resistance was not routinely measured, **Supplementary Fig. 1**, green/red horizontal bars). This temporal separation between training and testing data emulates forecasting resistance, as would be the case in real-life implementation of such a method. Area Under the Curve (AUC) of Receiver Operating Characteristic was used as a standard measure for predictive power³⁷.

Individually considering demographics, sample history and purchase history, we find that each of these sets of features had significant predictive power, yet their dominance varied across the different antibiotics (**Fig. 5**). For demographic parameters (Methods: Logistic regression “Demographics”), AUC ranged from barely 0.58 for ampicillin to values as high as 0.85 for amikacin, reflecting their weak and strong demographic dependencies, respectively (**Fig. 2b**). For sample history (Methods: Logistic regression “Sample history”), the predictive power was more consistent, with AUC varying between 0.64 for piperacillin/tazobactam to 0.83 for ertapenem. Purchase history (Methods: Logistic regression “Purchase history”) had barely any predictive power in drugs like meropenem and nitrofurantoin (AUC of 0.52 and 0.59, respectively), yet was higher for other drugs, especially for the fluoroquinolones (AUC of 0.78 for ofloxacin and 0.75 for ciprofloxacin). Comparing the relative strengths of these feature sets, we find that while they vary in dominance across the antibiotics, resistance to any given antibiotic was often well predicted by more than one set, suggesting that even higher predictability could be achieved by combining them.

Combining these feature sets, using logistic regression or GBDT models, much increased predictability of resistance. Using all features together in a complete logistic regression model (Methods: Logistic regression “Complete”), provided good to excellent predictability to all antibiotics (AUC ranged from 0.7 for amoxicillin/CA to and beyond 0.85 for ertapenem, amikacin and ofloxacin; **Fig. 5**). Predictability of resistance across all antibiotics was increased even further by the complete GBDT models (Methods; **Fig. 5**), most notably for meropenem, which was prone for overfitting (**Supplementary Fig. 8**). A closer look at the decision trees of the GBDT models identified the relative importance of the different features (**Supplementary Fig. 9**). While some features, like age and past resistance were important in all models, others like gender and time interval from last antibiotic purchase were more specific to each resistance. Together, these results demonstrate how machine learning models, trained on the combined dataset of demographics, sample history and drug purchase history, can provide high predictability of antibiotic resistance at the single-patient and single-infection levels.

Discussion

Analyzing a large longitudinal medical dataset, we demonstrate the predictability of antibiotic resistance in UTIs based on demographics, antibiotic resistance profile of past UTIs and purchase history of antibiotic drugs. These culture-free predictions could guide appropriate antibiotic prescription. Considering demographics, we found that - age, gender, pregnancy, and residence in a retirement home were strongly associated with resistance, showing complex even non-monotonic patterns specific to each of the different antibiotics. Utilizing repeated same-patient cultures in our database, we identified and characterized a personal component of “memory” of resistance, lasting for many months and even over a year and further contributing to predictability of resistance. Long memory was also observed between resistance and past drug purchases. Resistance to a given drug had long-lasting associations not only with past usage of this same drug, but also with other, even mechanistically unrelated, drugs. Yet, adjusting for linkage among resistances exposed direct selection where drug use leads specifically to its own cognate resistance at the single patient level. Taken together, these results are consistent with drug use directly selecting, at the single-patient level, for strains resistant to it and thereby selecting indirectly, likely through genetic linkage, to resistance to other antibiotics. Once resistance strains occur they persist in the body affecting the chance of resistant infections months ahead.

Some aspects of the data may complicate the interpretation of our results, especially wherever causality may be inferred. As purchase of a drug does not fully guarantee its

concurrent use, later usage of a purchased drug may bias our results towards higher odds ratio for purchases made long before infection. Conversely, we can not exclude that some patients have used antibiotics they did not purchase through MHS, which will bias our results towards lower odds ratio for drug purchases. Also, although culture data is routine for suspected UTIs, sending urine for a culture test is not obligatory. As a result, we assume some UTIs would be empirically treated without any culture record, and there could be enrichment for culture tests following treatment failure. Such enrichment towards cultures following treatment failure can generate bias towards measurements of more resistant samples, resulting in overestimation of the total frequency of resistance, especially for first-line treatment. It can further contribute to the strong short-term association of drug purchases with resistance, especially for first-line antibiotics. The level of enrichment for urine culture tests following treatment failure could itself depend on demographics, which can bias correlations of demographics with resistance. While we cannot exclude these biases, our analysis firmly demonstrates that, with all of these potential biases, resistance of urine samples can be well predicted based on the patient's specific demographics and clinical history.

The predictability of infection resistance profiles lays the basis for a future paradigm where clinicians will routinely consult machine learning algorithms for prescription of antibiotic treatment tailored by the patient records and their clinical history. While the key factors identified here can serve as the basis of such approach, we expect that the specific model, the exact coefficients and relative weights of predictors, must be tailored for each country or region. Indeed, these algorithms can also be dynamically and adaptively updated in real time as new data is acquired. We also expect that inclusion of additional patient specific factors, such as comorbidities, can further increase resistance predictability. In the longer term, these clinical-record based approaches could be integrated with genomics of the patient as well as of the pathogen^{38–44}. Implemented in the clinic, machine-learning guided personalized empirical prescription can reduce treatment failure as well as lower the overall use and misuse of antibiotics thereby assisting in the global effort of impeding the antibiotic resistance epidemic.

Methods

Data. Anonymized clinical records of urine culture tests (“culture reports”) and records of antibiotic purchases (“purchase reports”) were obtained from Maccabi Health Services (MHS) for the time period from July 2007 to June 2017. Random patient identifiers were used to link culture reports and antibiotic purchase reports.

Culture reports: Antibiotic resistance profiling of bacterial pathogens isolated from urine cultures was carried out centrally (in two locations until 2010, and in one central lab since). We retrieved 711,099 culture reports of positive samples from 315,047 patients total. Each report included: (1) Unique patient code; (2) Date of sample; (3) List of isolates cultured with species identification (typically one isolate per sample; 3.6% of samples had more than one isolate); (4) Resistance profile of the isolates from processed results of a VITEK 2 system given as Sensitive, Intermediate and Resistant for each drug tested. Not all antibiotics were measured for each sample, and only the 16 most commonly measured antibiotics were included for further analysis ($N_{Resistances} = 16$, **Supplementary Fig. 1, Table 1**). Additionally, for each of the 16 included antibiotic resistances, we removed time intervals where only a small percentage of samples were measured to avoid sampling bias (**Supplementary Fig. 1, gray**). (5) Demographics: age, gender, pregnancy of the patient, as well as a code for the reporting physician which we used to identify patients residence in retirement homes. (6) Referring lab identifier, which was used to assign patients to a city of residence.

Antibiotic purchase reports: All drug purchases by prescription are routinely recorded in MHS databases. We identified and retrieved all purchases by patients with culture reports by converting internal MHS drug codes to ATC classifications of antibiotics (**Supplementary Table 2**). Each purchase record included: (1) Unique patient code to be linked to the code of the culture record; (2) Internal MHS product code, which was translated to an ATC drug code, (3) Date of purchase.

Choice of drugs for analysis: We focused on 34 groups of antibiotic compounds ($N_{ATC} = 34$), including the 30 most purchased during the time of the study (groups 1-30, **Supplementary Table 2**) as well as less frequently purchased groups of compounds manually identified as cognate drugs of the 16 antibiotic resistances analysed (groups 31-33 in **Supplementary Table 2**).

Feature definition. For each cultured sample, we define the following parameters used for the logistic regression and the gradient boosting decision trees:

Sample resistance profile: For each urine sample, we define Y_k as 0 for sensitive and intermediate and 1 for resistant to antibiotic k ($1 \leq k \leq N_{Resistances}$). If the sample had multiple isolates, Y_k was assigned 1 if at least one isolate was resistant. Missing resistance measurements are defined as N/A, and for each antibiotic k only samples which have defined resistance to it are used when training or testing its Logistic Regression or Gradient Boosting Decision Trees (GBDT).

Demographics: X^{Gender} : 0/1 for males/females; $X^{Pregnancy}$: 0/1 indicating pregnancy; $X^{Ret.Home}$: 0/1 indicating residence in retirement homes; X^{Age} : patient age in years at time of UTI sampling; X^{Date} : date of sample in days starting at 1-July-2007; X^{Season} : the day of sample within the calendar year, where 1 is 1-Jan and 365 is 31-Dec; X_j^{City} : 0/1 indicating residence in city j ($1 \leq j \leq N_{City}$, $N_{City} = 20$, Supplementary Fig. 2 for City list; we use Tel-Aviv as a reference); for GBDT, nominal $X^{City} \in \{1, \dots, 20\}$ was used instead.

Sample history: For a given sample, we consider all earlier samples taken from the same patient (if any). We bin the time difference between any such earlier sample and the current sample, $t = t_{Past\ sample} - t_{Sample}$ (t is negative, designating past events), into one of 8 logarithmically spaced time bins ($i = 1, 2, \dots, N_{time\ bins}$, $N_{time\ bins} = 8$, a bin i is defined by $t_i \leq t < t_{i-1}$, where the boundaries of these time bins are $t_0 = 0$, $t_1 = -1$, $t_2 = -2$, $t_3 = -4, \dots, t_8 = -128$ weeks). We then calculated $X_{i,k}^{Previous\ Resist}$ and $X_{i,k}^{Previous\ Sensitive}$ as the number of prior cultures within time bin i , with sample resistance Y^k equal to 0 or 1 (Resistant or Sensitive), respectively.

Drug purchase history: For each urine sample, we consider all earlier drug purchases made by the same patient. We bin the time difference between the urine sample date and a given past purchase, $t = t_{Purchase} - t_{Sample}$, into 8 time bins as above. For each sample, we then calculate X_{ij}^{ATC} as the number of purchases of drugs of a given group j ($1 \leq j \leq N_{ATC}$, **Supplementary Table 2**) made by the patient during time bin i . For the GBDT model, we also included as additional features the types j of the 3 latest purchases (nominal features) and their purchase-to-sample time difference (t , in days)

Cross-resistance: To resolve direct versus indirect associations of drug purchase and resistance, we adjusted the logistic regression of resistance to a given antibiotic k as a function of past drug purchases by the resistances to all other drugs j which are non-analogous to k . We define $A_{k,j}$ as a binary variable equals 0 and 1 for analogous versus non-analogous drug pairs, respectively. “Analogous” pairs are defined as antibiotics which have exceptionally high cross-resistance ($A_{k,j} = 0$ for

$corr(Y_k, Y_j) > A_{threshold}$; we use $A_{threshold} = 0.7$ which corresponds to drug pairs of the same class; see pairs labeled with 'x' in **Supplementary Fig. 7**). We then add as features in the regression analysis of a given antibiotic k the resistance measurements Y_j to all antibiotics j for which $A_{k,j} = 1$. Note: These cross-resistance features provide information from the focal sample and were used only in the analysis of direct/indirect effect of purchases (**Fig. 4** and **Supplementary Fig. 6c**). Importantly, they were not used when evaluating predictability of resistance (**Fig. 5**).

Logistic regression. Logistic regression of resistance for each antibiotic was performed via the Matlab glmfit function. For each of the 16 antibiotics k , the probability of resistance P_k was fit to the sample resistance Y_k for all urine samples which had measurement of resistance to k either across the entire 10 year dataset (for **Figs. 2,4**), or across the “training period” (for the analysis of predictive power of **Fig. 5**; see **Supplementary Fig. 1** for definition of the training period for each of the 16 antibiotics). The different logistic models included combinations of the following 10 terms:

$ln\left(\frac{P_k}{1-P_k}\right) = C_k^{Const} +$	Term #
$C_k^{Gender} X^{Gender} +$	#1
$C_k^{Pregnancy} X^{Pregnancy} +$	#2
$C_k^{Ret.Home} X^{Ret.Home} +$	#3
$\sum_{j=1}^{N_{City}} C_{k,j}^{City} X_j^{City} +$	#4
$C_k^{Age_1} \left(\frac{X^{Age}}{100}\right) + C_k^{Age_2} \left(\frac{X^{Age}}{100}\right)^2 +$	#5
$C_k^{Date_1} \left(\frac{X^{Date}}{365 \cdot 10}\right) + C_k^{Date_2} \left(\frac{X^{Date}}{365 \cdot 10}\right)^2 +$	#6
$C_k^{Season_1} \sin\left(2\pi \frac{X^{Season}}{365}\right) + C_k^{Season_2} \cos\left(2\pi \frac{X^{Season}}{365}\right) +$	#7
$\sum_{i=1}^{N_{time\ bins}} C_{k,i}^{Previous\ Resist} X_{k,i}^{Previous\ Resist} + C_{k,i}^{Previous\ Sensitive} X_{k,i}^{Previous\ Sensitive} +$	#8
$\sum_{i=1}^{N_{time\ bins}} \sum_{j=1}^{N_{ATC}} C_{k,i,j}^{ATC} X_{i,j}^{ATC} +$	#9
$\sum_{j=1}^{N_{Resistances}} A_{k,j} C_{k,j}^{Cross\ Resist} Y_j$	#10

Different combination of the above terms were used in the different regression models as follows (each row in the Table represents a logistic model that was applied to each of the 16 antibiotics):

Display item	Model name	Regression terms	Fit data
Fig. 2b Sup. Table 1	Demographics	#1-#7	All data
Sup. Table 1	Gender, unadjusted	#1	All data
Sup. Table 1	Pregnancy, adjusted for gender	#2 and #1	All data
Sup. Table 1	Ret.Home, unadjusted	#3	All data
Sup. Table 1	City, unadjusted	#4	All data
Sup. Table 1	Age, unadjusted	#5	All data
Sup. Table 1	Date, unadjusted	#6	All data
Sup. Table 1	Season, unadjusted	#7	All data
Fig. 4a,c,d Sup Fig. 6a	Purchase history	#9	All data
Sup Fig. 6b	Purchase history, adjusted for demographics	#9 and #1-#7	All data
Fig. 4b,c,d Sup Fig. 6c	Purchase history, adjusted for cross-resistance	#9 and #10	All data
Fig. 5	Demographics	#1-#7	Training range*
Fig. 5	Sample history	#8	Training range*
Fig. 5	Purchase history	#9	Training range*
Fig. 5	Complete	#1-#9	Training range*

* See Supplementary Fig. 1 for training range of each of the 16 resistances.

Calculating odds ratios from logistic regression. For each antibiotic k , odds ratios were calculated from the coefficients of above logistic regressions.

Binary variables: For the binary variables Gender, Pregnancy and Retirement Home, odds ratio were defined as: $OR_k^{Gender} = \exp(C_k^{Gender})$ female versus male, $OR_k^{Pregnancy} = \exp(C_k^{Pregnancy})$ pregnant versus non-pregnant, $OR_k^{Ret.Home} = \exp(C_k^{Ret.Home})$ retirement home residence versus patients not residing in retirement homes.

Categorical variables: For City, which is a categorical variable, odds ratios for each city j relative to the reference (Tel-Aviv) is given by $OR_k^{City} = \exp(C_{k,j}^{City})$, where $C_{k,j}^{City}$ are reported in **Supplementary Table 1**. In **Fig. 2**, we report for City the odds ratio $OR_k^{City\ max} = \exp(C_{k,j_{max}}^{City})$, where j_{max} is the index of the city with largest effect, accounting for error (j_{max} is defined for each antibiotic k as the city j with maximal $|C_{k,j}^{City}| - \Delta C_{k,j}^{City}$, where $\Delta C_{k,j}^{City}$ are the standard errors for the regression coefficients $C_{k,j}^{City}$).

Oscillatory variables: For seasonality, the regression coefficients $C_k^{Season_1}$ and $C_k^{Season_2}$ and their CIs are given **Supplementary Table 1**. In **Fig. 2** we provide the odds ratios of the oscillation amplitude: $OR_k^{Season} = \exp\left(\sqrt{(C_k^{Season_1})^2 + (C_k^{Season_2})^2}\right)$.

Quadratic variables: For the quadratic variables Age and Date, we again report all the individual regression coefficients and their CIs in **Supplementary Table 1**. In **Fig. 2b**, we report, for each antibiotic k , effective odds ratios defined as the ratios between the maximal and minimal expected odds taken across the relevant range of these continuous variable ($0 \leq X^{Age} \leq 100$ and $0 \leq X^{Date} \leq 10 \cdot 365$):

$$OR_k^{Age} = \exp\left[\max_{0 \leq x \leq 1} (C_k^{Age_1} x + C_k^{Age_2} x^2) - \min_{0 \leq x \leq 1} (C_k^{Age_1} x + C_k^{Age_2} x^2)\right]$$

$$OR_k^{Date} = \exp\left[\max_{0 \leq x \leq 1} (C_k^{Date_1} x + C_k^{Date_2} x^2) - \min_{0 \leq x \leq 1} (C_k^{Date_1} x + C_k^{Date_2} x^2)\right]$$

Note that when these quadratic dependencies are monotonic within the relevant range ($0 \leq x \leq 1$), the above formulas become simply: $OR_k^{Age} = \exp(C_k^{Age_1} + C_k^{Age_2})$,

$$OR_k^{Date} = \exp(C_k^{Date_1} + C_k^{Date_2}).$$

Analysis of “memory” across sample pairs. To analyze “memory” of resistance across samples, we considered all pairs of samples from the same patient (across all patients with 2-10 samples) and binned them according to their time difference $t = t_1 - t_2$ (where t_1 and t_2 are the sample dates of the early and late sample; t is always negative, indicating information on current sample from past samples) into time bins as indicated by the bars in Fig. 3. In each time bin and for each antibiotic, we counted $N_{R \rightarrow R}$, $N_{R \rightarrow S}$, $N_{S \rightarrow R}$, and $N_{S \rightarrow S}$ as the number of urine sample pairs where the early and late samples are Resistant, or Sensitive (for example $N_{R \rightarrow S}$ is the number of same-patient sample pairs, within the time difference bin, where first sample is Resistant and the second Sensitive to the given focal antibiotic. For any antibiotic, only samples for which resistance to the antibiotic was measured were considered). We then calculated for each time difference bin the resistance odds ratio: $OR_{pairs} = (N_{R \rightarrow R} / N_{R \rightarrow S}) / (N_{S \rightarrow R} / N_{S \rightarrow S})$.

Gradient Boosting Decision Trees (GBDT). GBDT is an ensemble method combining regression trees with weak individual predictive performances, into a single high-performance model. This is done by iteratively fitting decision trees, each iteration targeting the prediction residuals of the preceding tree. The final model is built by combining weighted individual tree contributions, with weights proportional to their performances. For each of the 16 antibiotics, a boosted decision tree ensemble was fitted using all features as defined above (demographics, sample history and drug purchase history) on the training set as defined by the training time period (**Supplementary Fig. 1**, green bars). This training dataset was sampled to balance resistant/sensitive label frequency. For parameter tuning, a validation dataset was sampled from the training set to be used for model selection (20%). For the estimator of the i^{th} iteration, a decreasing learning rate η_i was used such that $\eta_i = \eta_0 \alpha^i$, with an annealing rate $\alpha = 0.99$ and an initial learning rate $\eta_0 = 0.1$. To further promote a diverse ensemble of individual estimators, a 0.9 feature-sampling and observation-sampling rates were used. Fitting of interaction effects is controlled by varying the size of the individual regression trees, with tree estimator of depth k producing models with up to k -way interactions. The model was tuned to match data complexity by iteratively increasing tree depth limit of all ensemble estimators while evaluating performance on the validation set, selecting the best depth for each antibiotic.

Ethical approval. The study protocol was approved by the ethics committee of Assuta Medical Center.

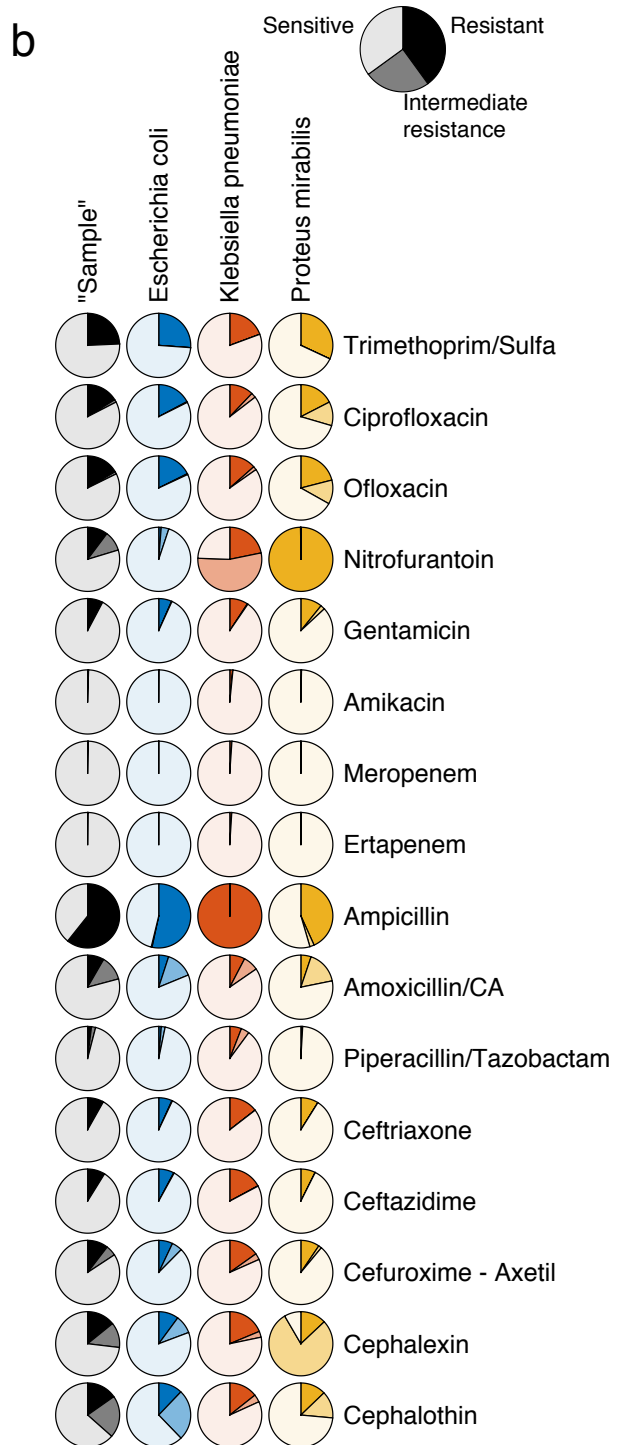
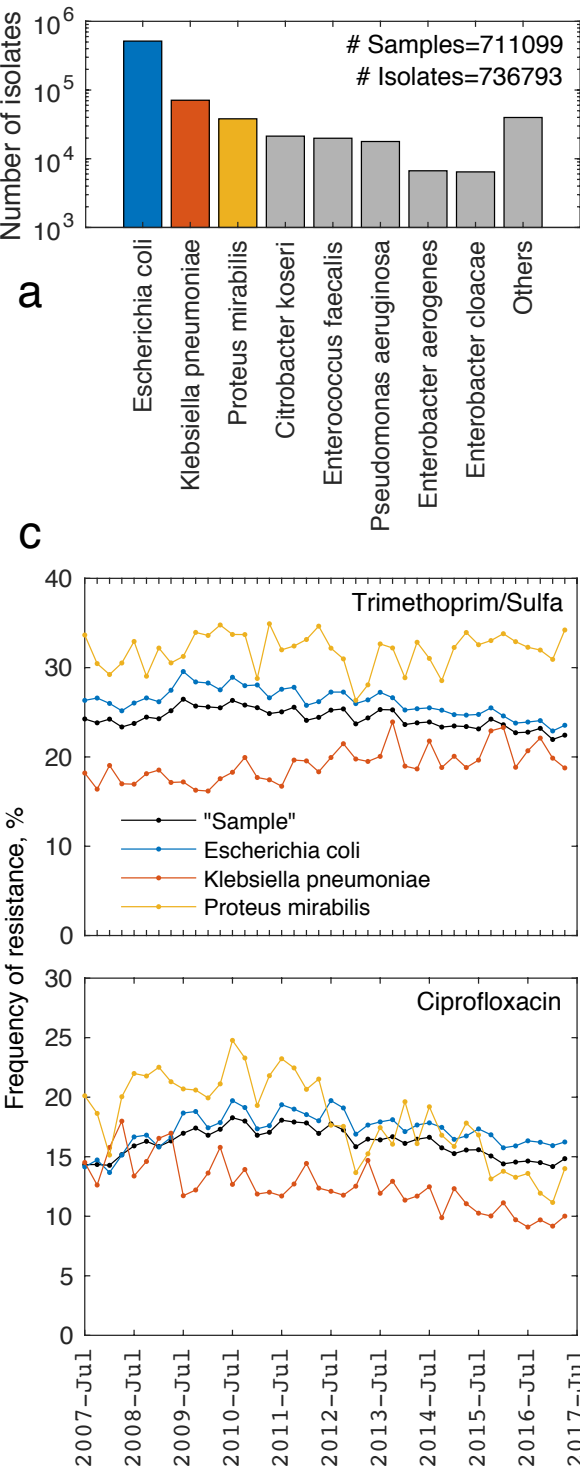
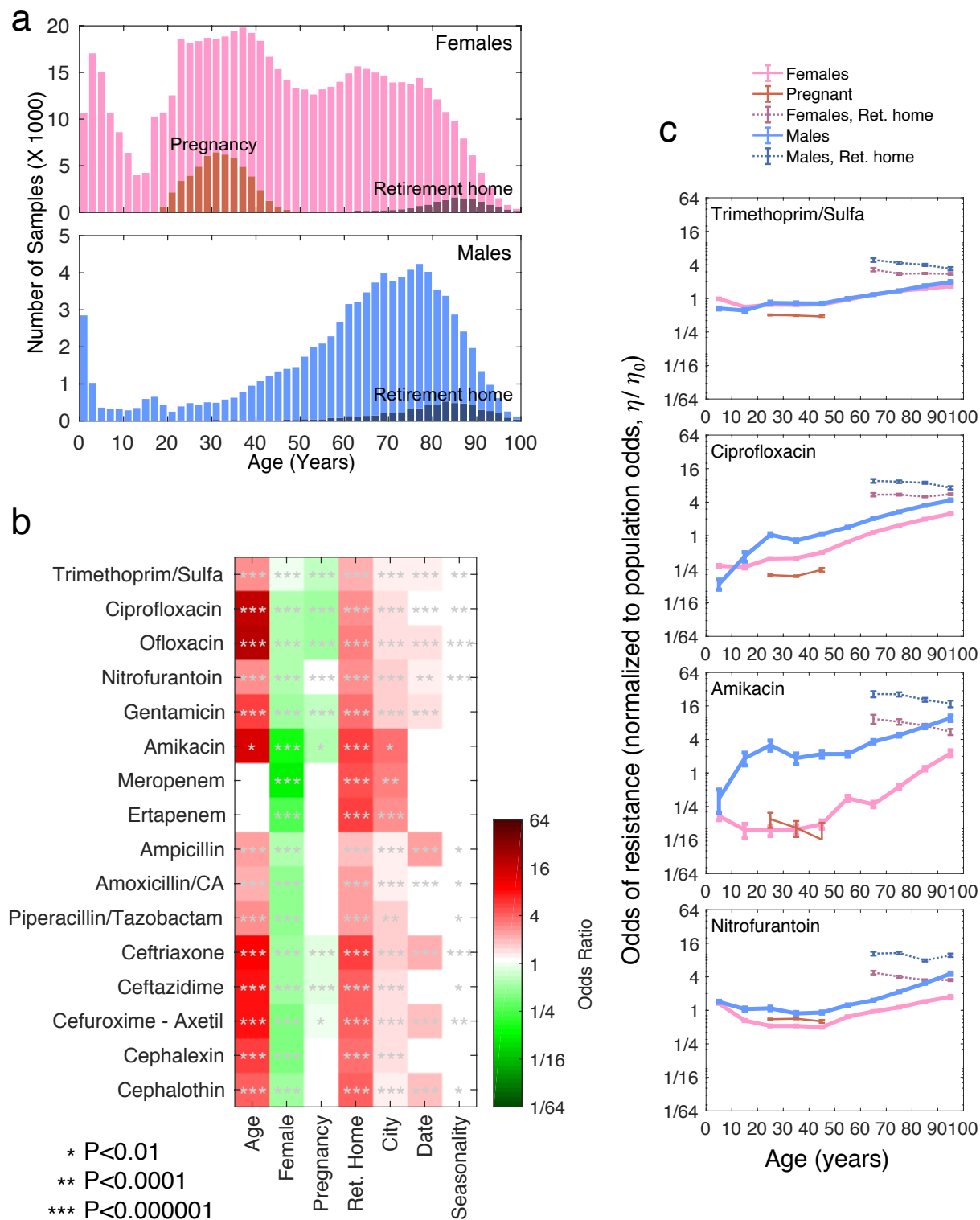


Figure 1: Frequency of bacterial species and antibiotic resistance in urinary tract infections.



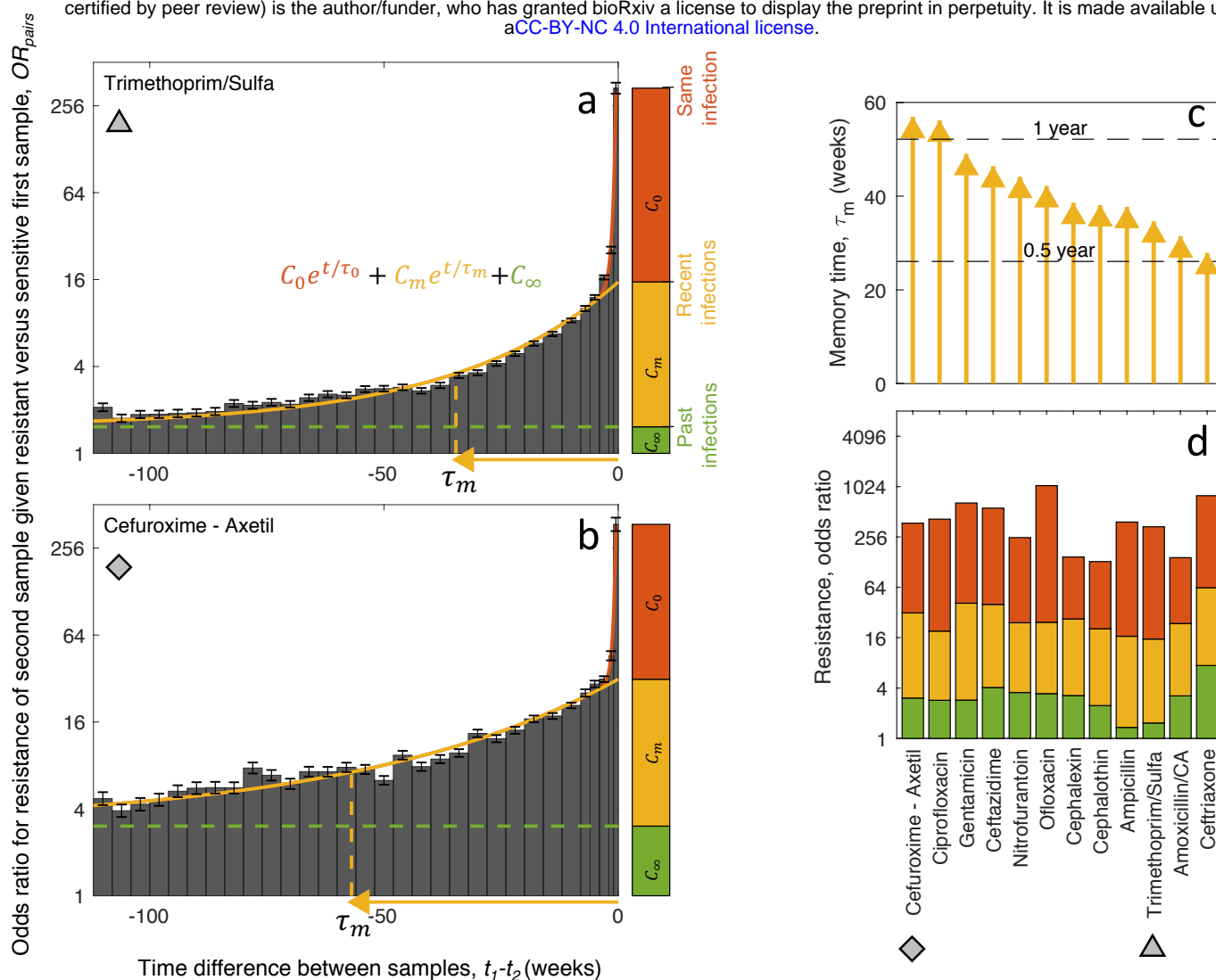


Figure 3: Long term “memory” of resistance across same-patient samples.

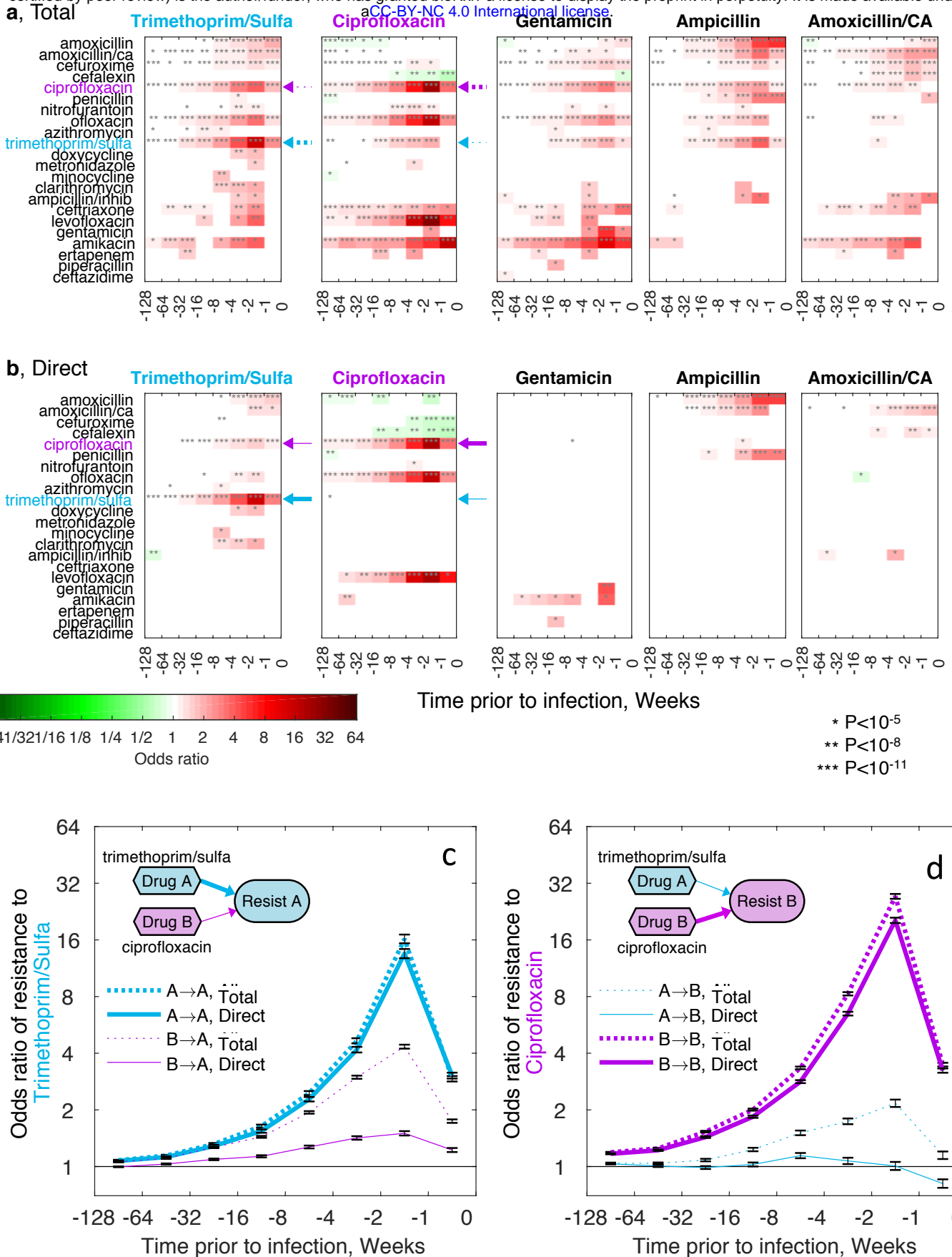


Figure 4: Past purchase of a given antibiotic has long-term direct association with its cognate resistance, which leads, through inherent linkage among resistances, also to indirect associations with other resistances.

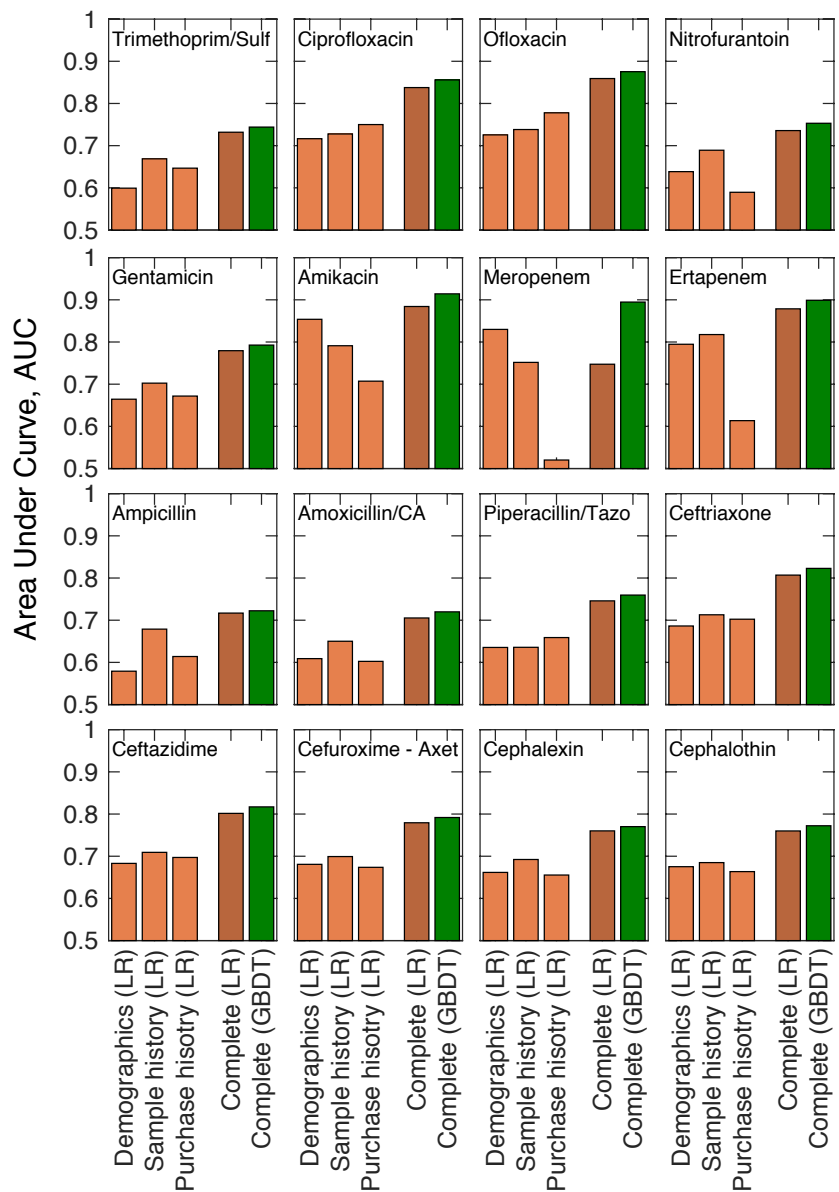


Figure 5: Prediction of resistance at the single-infection level based on patient demographics, urine sample history and drug purchase history.

Figure Captions

Figure 1: Frequency of bacterial species and antibiotic resistance in urinary tract infections. (a) Species abundance across the entire UTI dataset (July 2007-June 2017, 711099 samples, 736793 tested isolates). (b) The frequency of resistance and intermediate resistance to the 16 most frequently measured antibiotic drugs for the three most common bacterial species and for the urine sample as a whole (dark to light shades represent Resistant, Intermediate and Sensitive, respectively; Sample resistance is defined, for each urine sample as the highest resistance across all its isolates). (c) Frequencies of resistance for each of the three common species (colored lines) and the sample resistance (black lines) over the 10 year sampling time, for two representative antibiotics: trimethoprim/sulfa (top) and ciprofloxacin (bottom; See **Supplementary Fig. 2** for all antibiotics.). Data points represent quarterly averages.

Figure 2: Strong and antibiotic-specific associations of resistance with demographic factors. (a) Distribution of positive urine cultures across major demographic factors: Age, Gender (top: Females, bottom: Males), Pregnancy (red) and Retirement home residence (black). (b) Adjusted odds ratios of resistance for each demographic variable (Methods: Logistic regression “Demographics”; for all adjusted and unadjusted regression coefficients see **Supplementary Table 1**). Asterisks indicate statistical significance and insignificant coefficients are shown as blank (white shade, $P > 0.01$). (c) Odds of resistance as a function of age showing qualitatively distinct patterns for four representative antibiotics (Odds η are normalized to the average odds η_0 across the population) UTI samples are separated into five non-overlapping categories: men not residing in retirement homes (blue), men residing in retirement homes (dotted blue), women not pregnant and not residing in retirement homes (magenta), women in retirement home (magenta dotted), and pregnant women (red). See **Supplementary Fig. 4** for all antibiotics.

Figure 3: Long term “memory” of resistance across same-patient samples. (a,b) Odds ratio of resistance of a urine sample given resistant versus sensitive earlier sample from the same patient, as a function of the time difference between the two samples, for trimethoprim/sulfa and cefuroxime-axetil ($OR_{pairs} = (N_{R \rightarrow R}/N_{R \rightarrow S}) / (N_{S \rightarrow R}/N_{S \rightarrow S})$, Methods; **Supplementary Fig. 5** for all antibiotics). Odds ratios are well fitted with $OR_{pairs} = C_0 e^{t/\tau_0} + C_m e^{t/\tau_m} + C_\infty$, representing short and long memory time scales and a timeless constant (red, yellow and green respectively). The magnitudes of these terms are shown as stacked bars on the right and the longer half life (τ_m) is indicated across the time axis (yellow arrow). (c) Time scale of the long-term memory of resistance τ_m for different antibiotics. (d) The

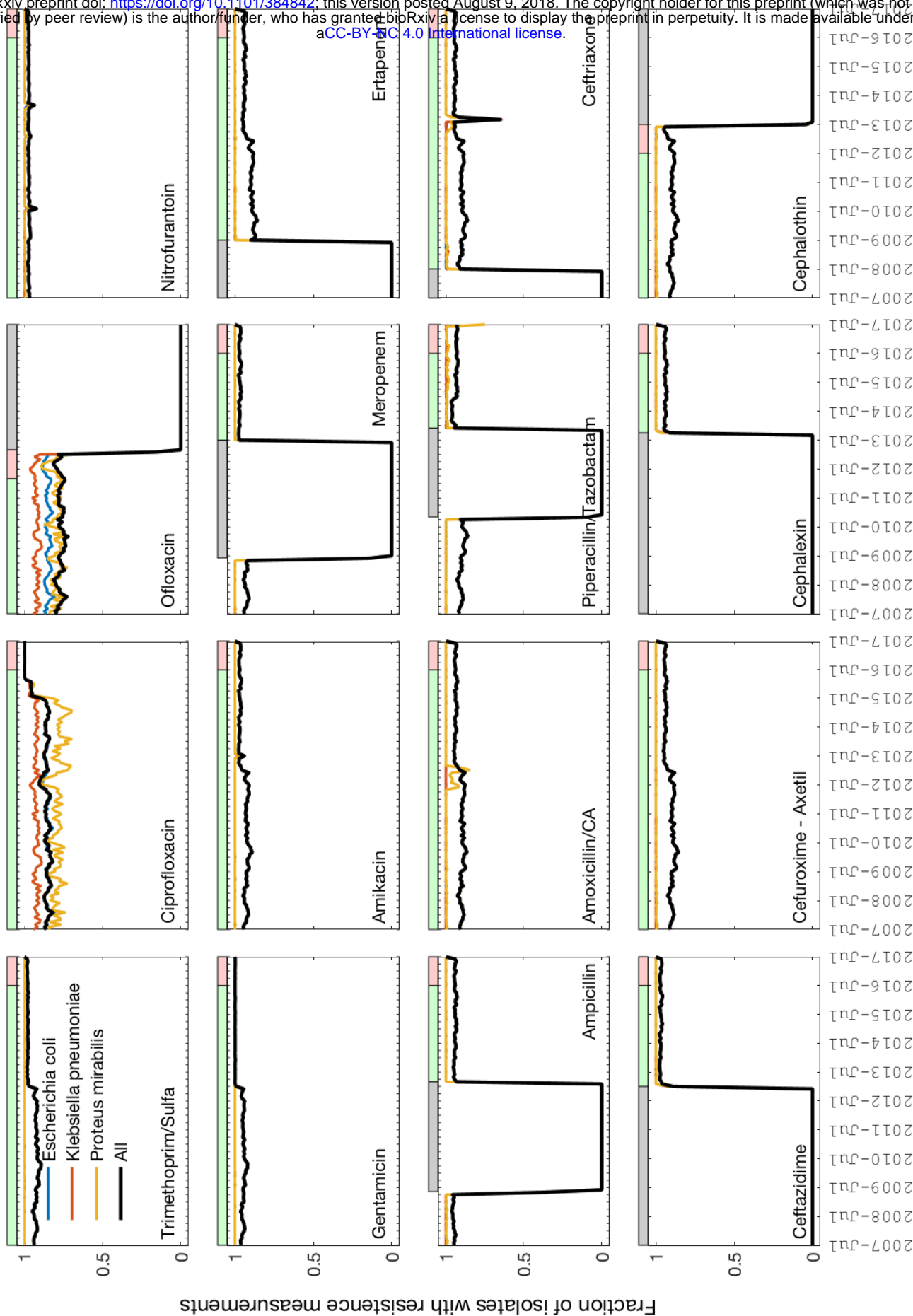
magnitude of short, long and timeless memory for different antibiotics (red, yellow, green bars).

Figure 4: Past purchase of a given antibiotic has long-term direct association with its cognate resistance, which leads, through inherent linkage among resistances, also to indirect associations with other resistances. (a) Multivariate logistic regression models for association of resistance to 5 representative antibiotics (for all antibiotics see **Supplementary Fig. 6a**) with past purchases of different drugs at different time intervals prior to the time of sample (Methods: Logistic regression “Purchase history”; Values represent odds ratios for a single purchase of a specific drug at a specific time interval). Resistance shows long term association with past purchase of its matching antibiotic (cognate, see examples indicated by thick dashed arrows for trimethoprim/sulfa and ciprofloxacin) as well as with many other antibiotics (noncognate, see examples indicated by thin dashed arrows). (b) Logistic regression model as in (a) but, for each focal antibiotic A, adjusting for resistance of each urine sample to antibiotics different than A (Methods: Logistic regression “Purchase history adjusted for cross-resistance”; for all antibiotics see **Supplementary Fig. 6c**). This adjustment disentangles indirect effect where resistance to drug A is associated with purchase of drug B due to inherent association of resistance to B with resistance to A. This adjusted model is seen to diminish or even completely remove noncognate drug-to-resistance associations while mostly leaving the cognate associations un-attenuated. (c,d) Association of trimethoprim/sulfa and ciprofloxacin with purchases of their cognate drugs (cyan and magenta, respectively). Compare difference between total (Methods: Logistic regression “Purchase history”, dashed lines) and direct effect (Methods: Logistic regression “Purchase history adjusted for cross-resistance”, solid lines) for cognate (thick lines) versus noncognate (thin lines) drugs.

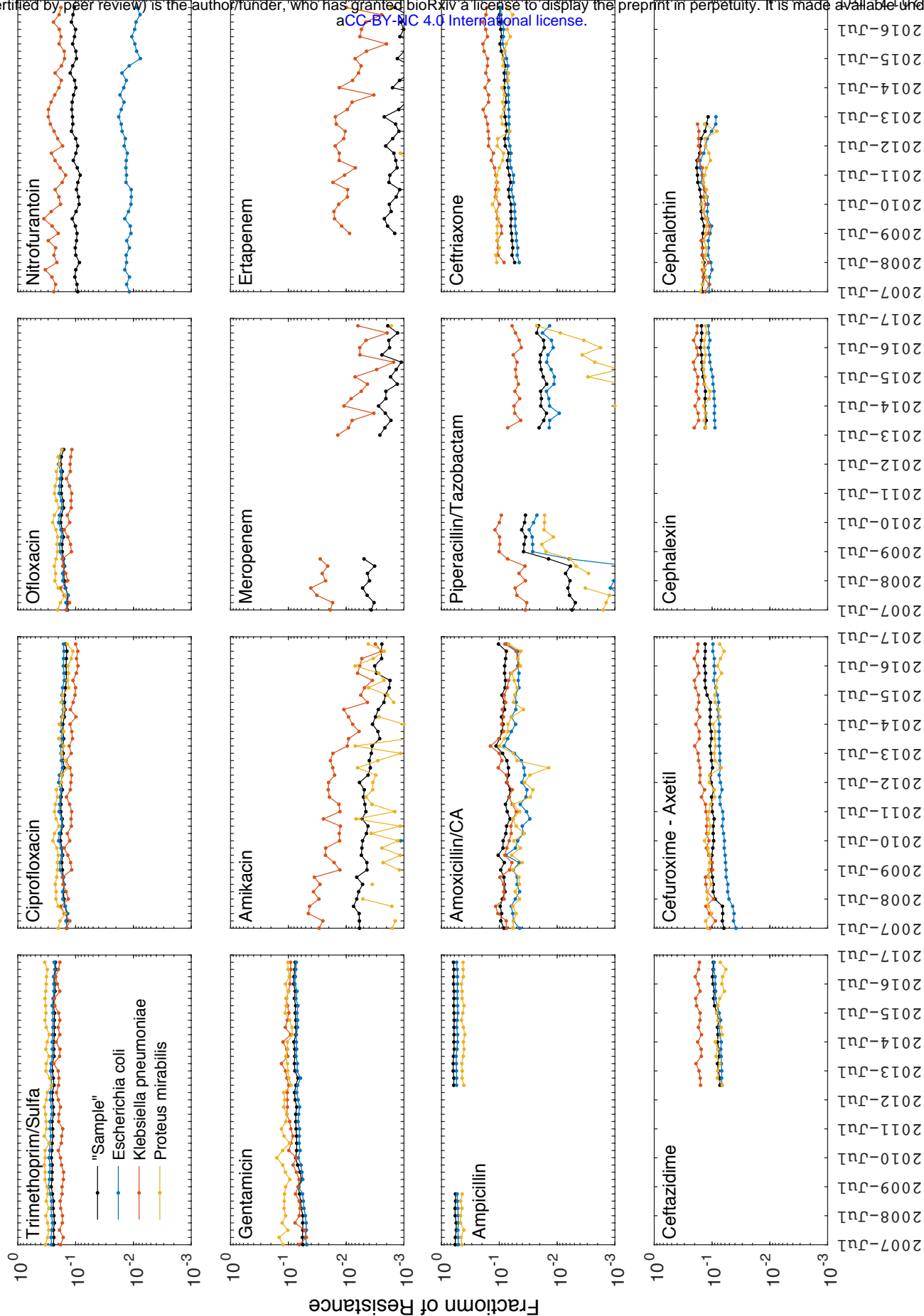
Figure 5: Prediction of resistance at the single-infection level based on patient demographics, urine sample history and drug purchase history. Area Under the Curve (AUC) of Receiver Operator Characteristic for prediction of resistance to each antibiotic using demographics, sample history and purchase history individually in a Logistic Regression model (LR, light orange), as well as when all of these feature sets are combined together in LR (dark orange) or in Gradient Boosting Decision Trees (GBDT, green). Models were trained on urine samples within the training date range (typically 9 years; see **Supplementary Fig. 1**) and tested on the urine samples collected during the year following the training period (green and red horizontal bars, **Supplementary Fig. 1**).

Table 1: List of antibiotic resistances analyzed in the study

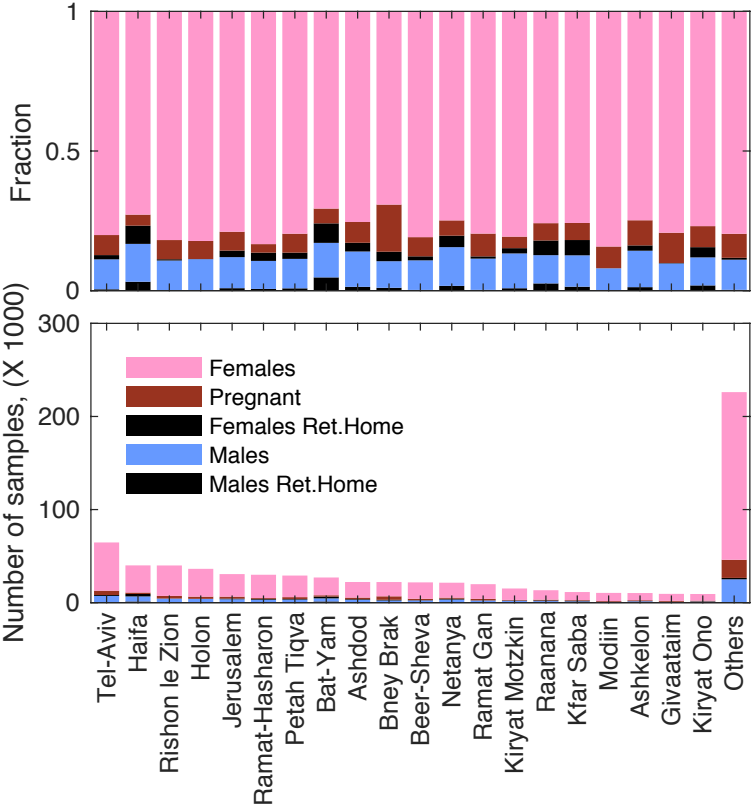
Antibiotics	Class
Trimethoprim/Sulfa	DHFR inhibitor
Ciprofloxacin	fluoroquinolones
Ofloxacin	
Nitrofurantoin	nitrofurans
Gentamicin	aminoglycosides
Amikacin	
Meropenem	carbapenems
Ertapenem	
Ampicillin	penicillins (+/- β -lactamase inhibitors)
Amoxicillin/CA	
Piperacillin/Tazobactam	
Ceftriaxone	cephalosporins
Ceftazidime	
Cefuroxime – Axetil	
Cephalexin	
Cephalothin	



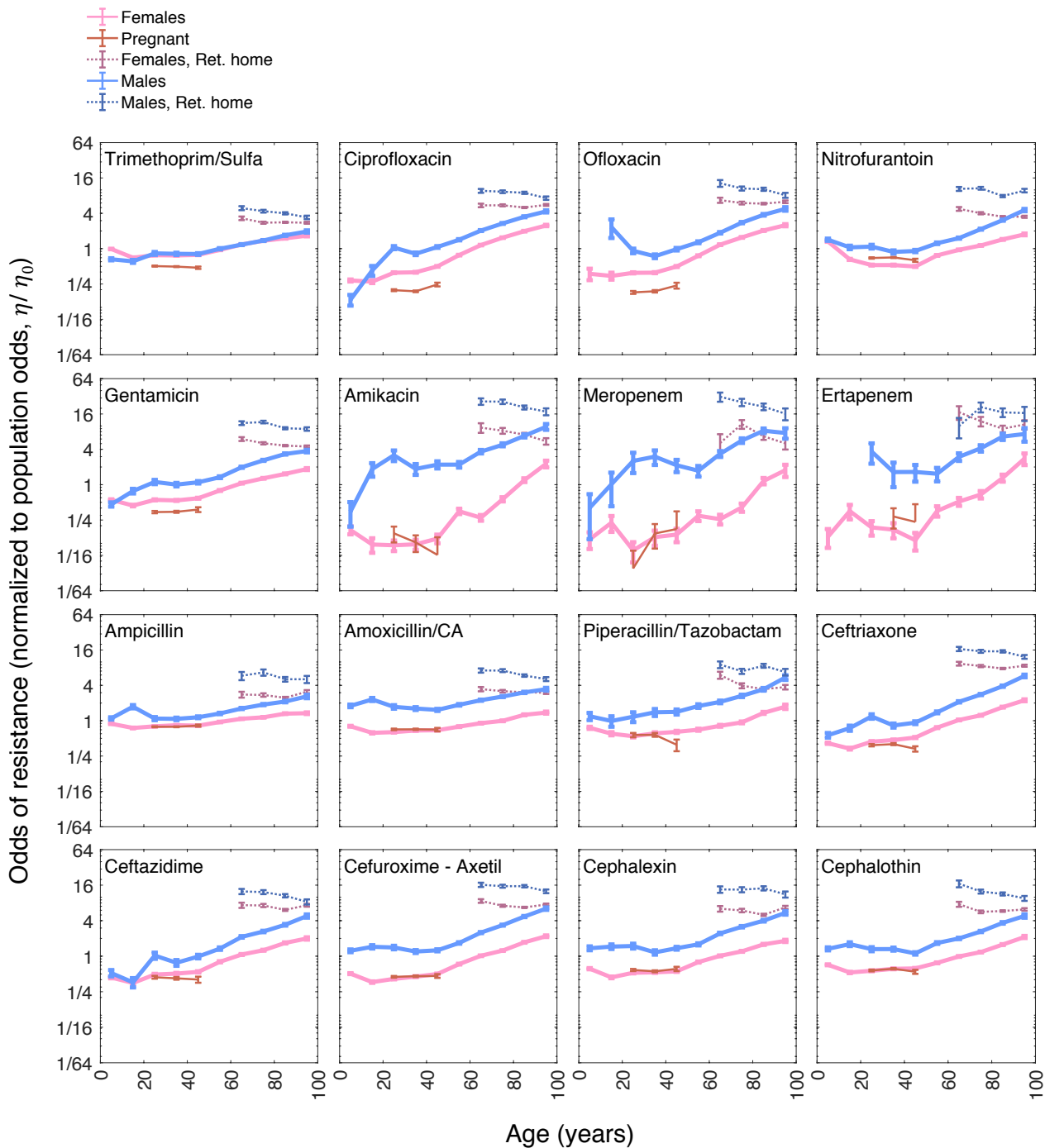
Supplementary Figure 1: Availability of resistance measurements over time.



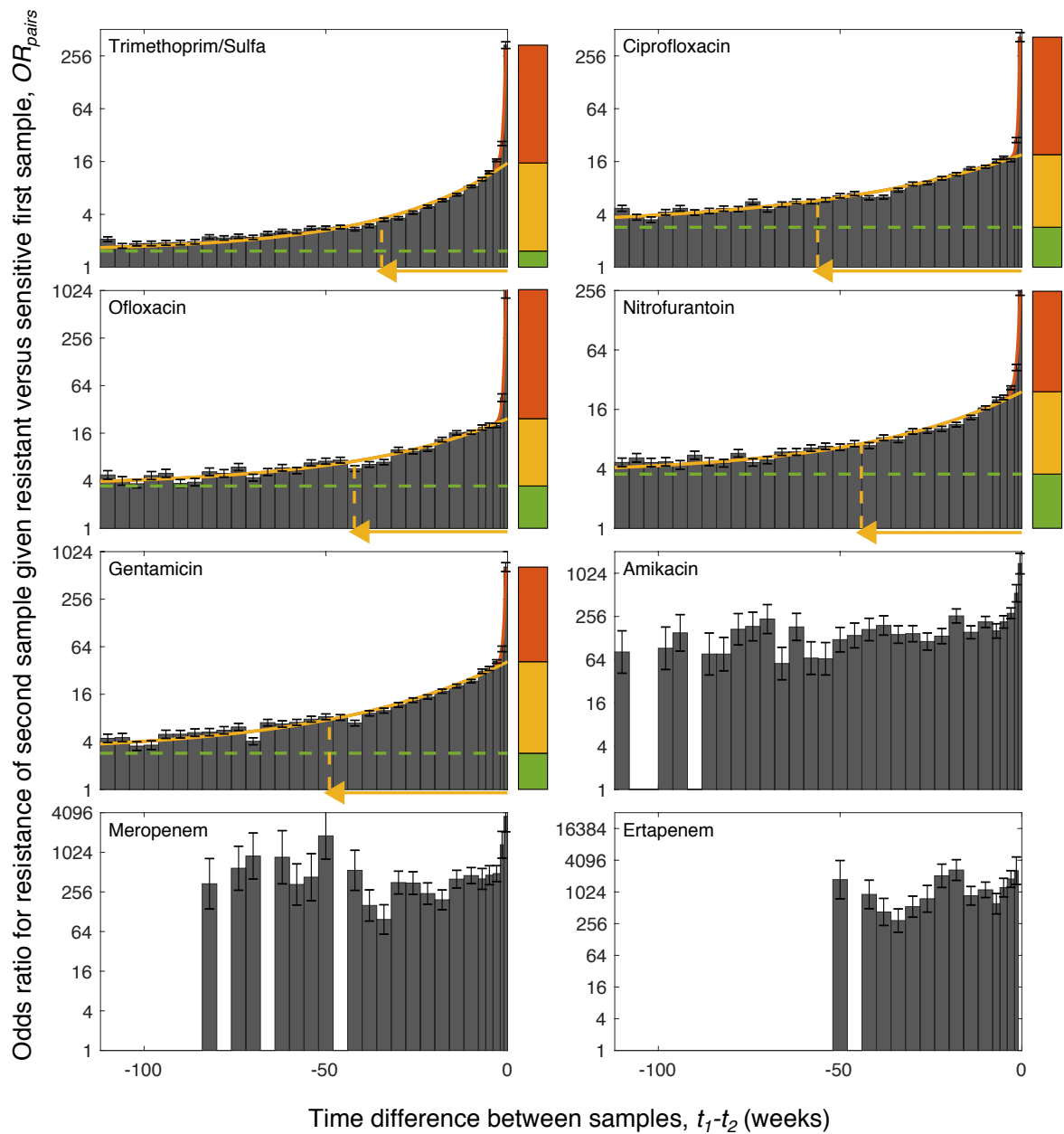
Supplementary Figure 2: Frequency of resistance over time.



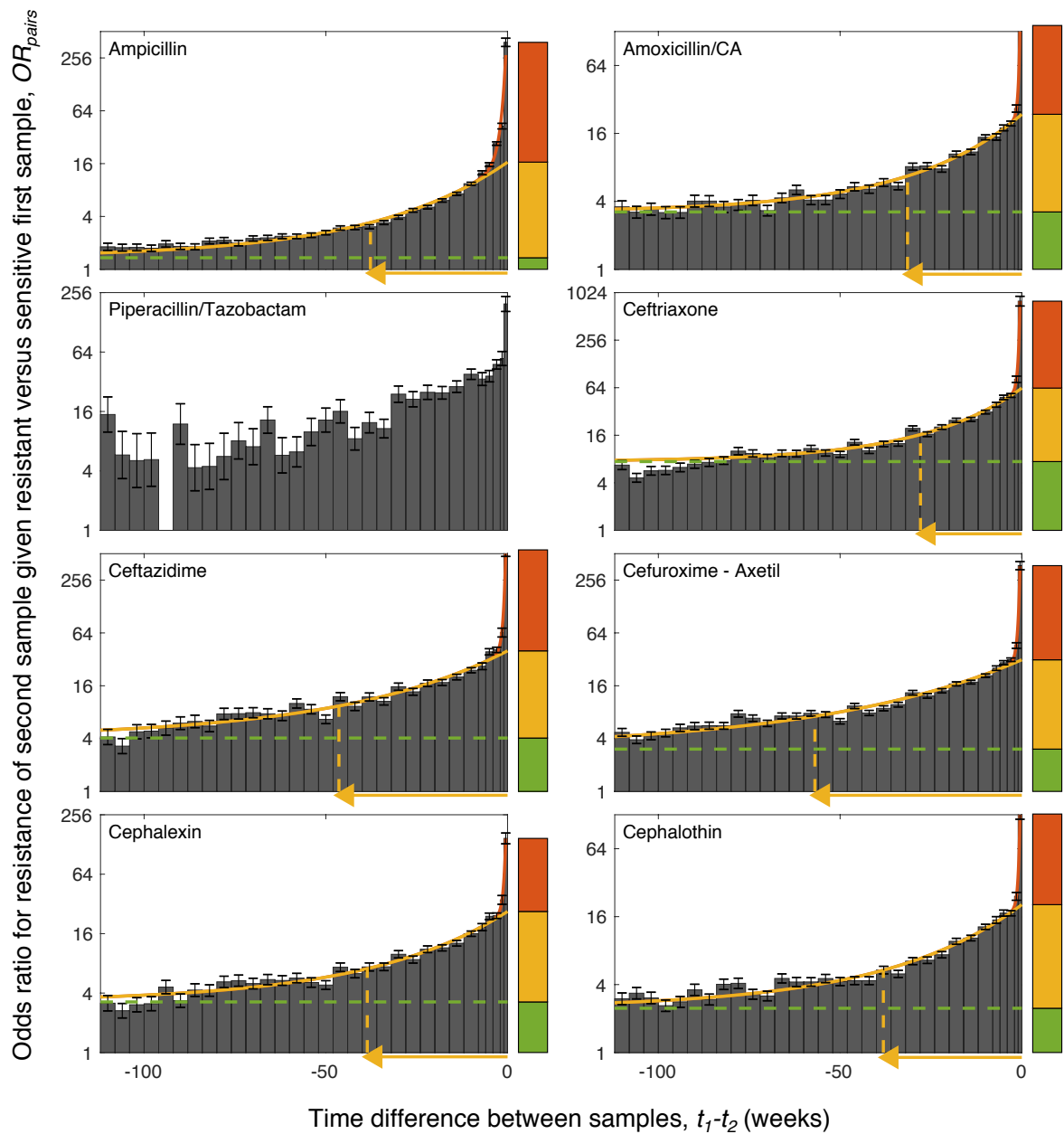
Supplementary Figure 3: Distribution of UTI demographics by city.



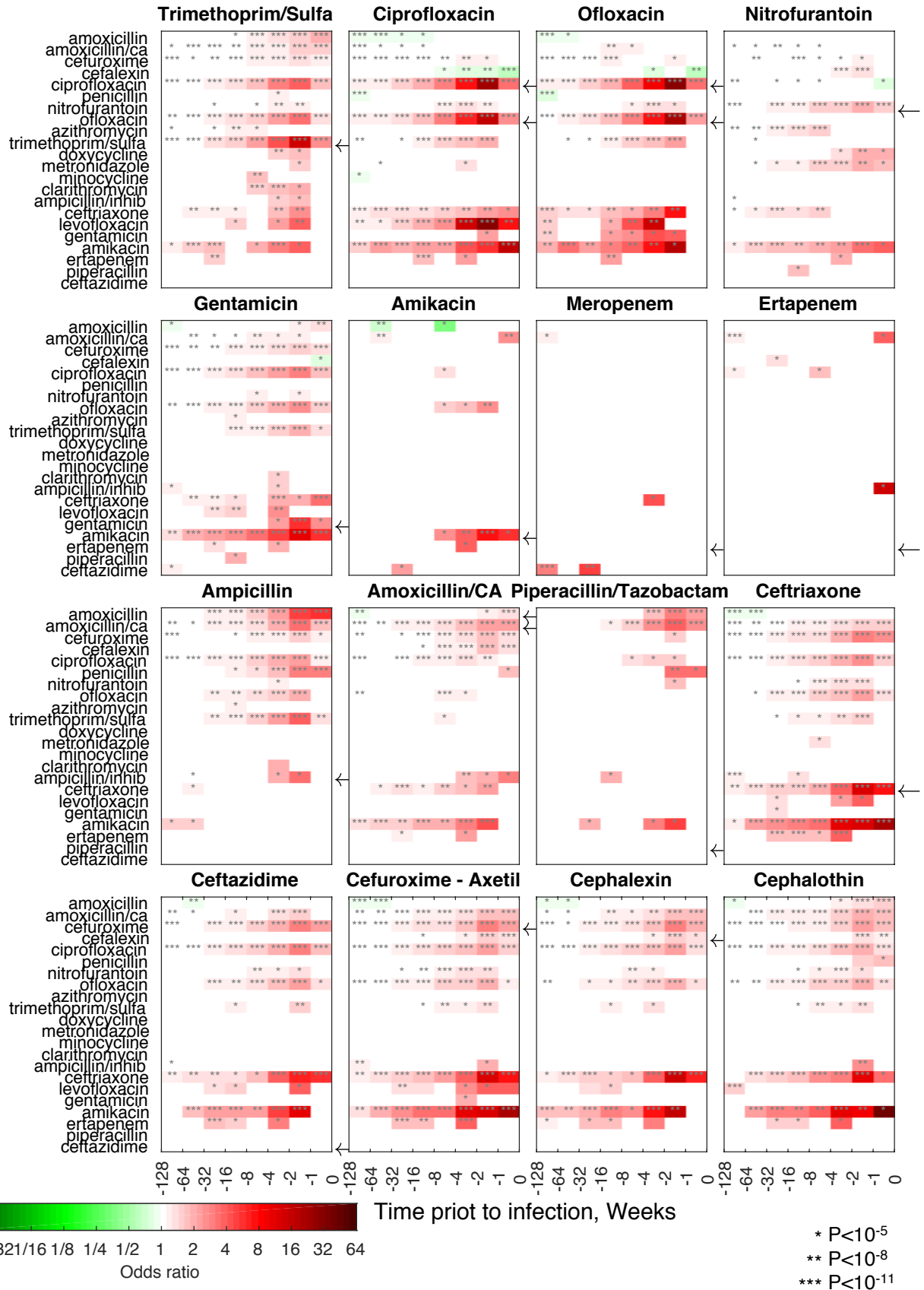
Supplementary Figure 4: Odds of resistance as a function of age for different demographic groups.



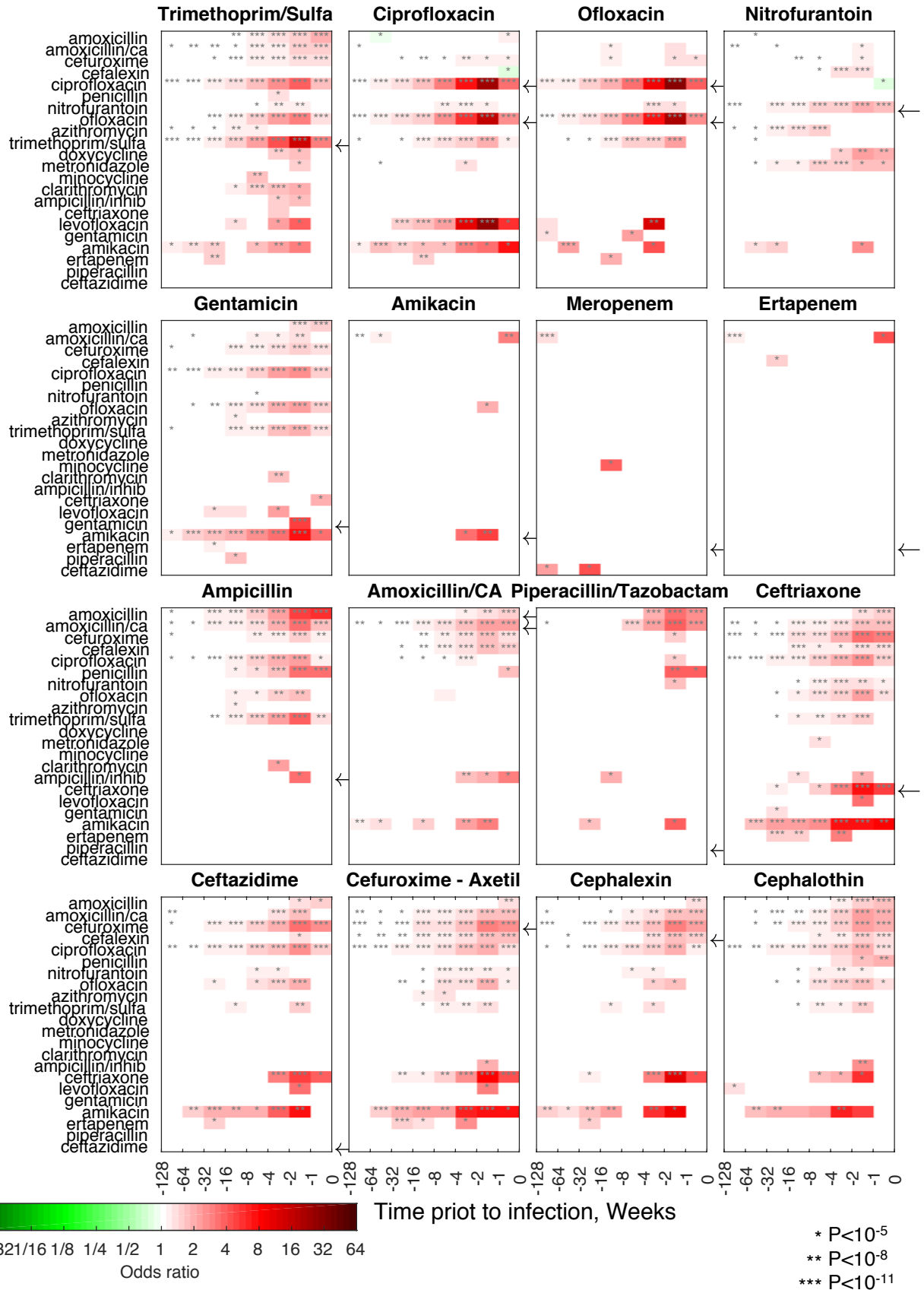
Supplementary Figure 5: Odds ratio of resistance across same-patient sample pairs.



Supplementary Figure 5: Odds ratio of resistance across same-patient sample pairs. (cont.)



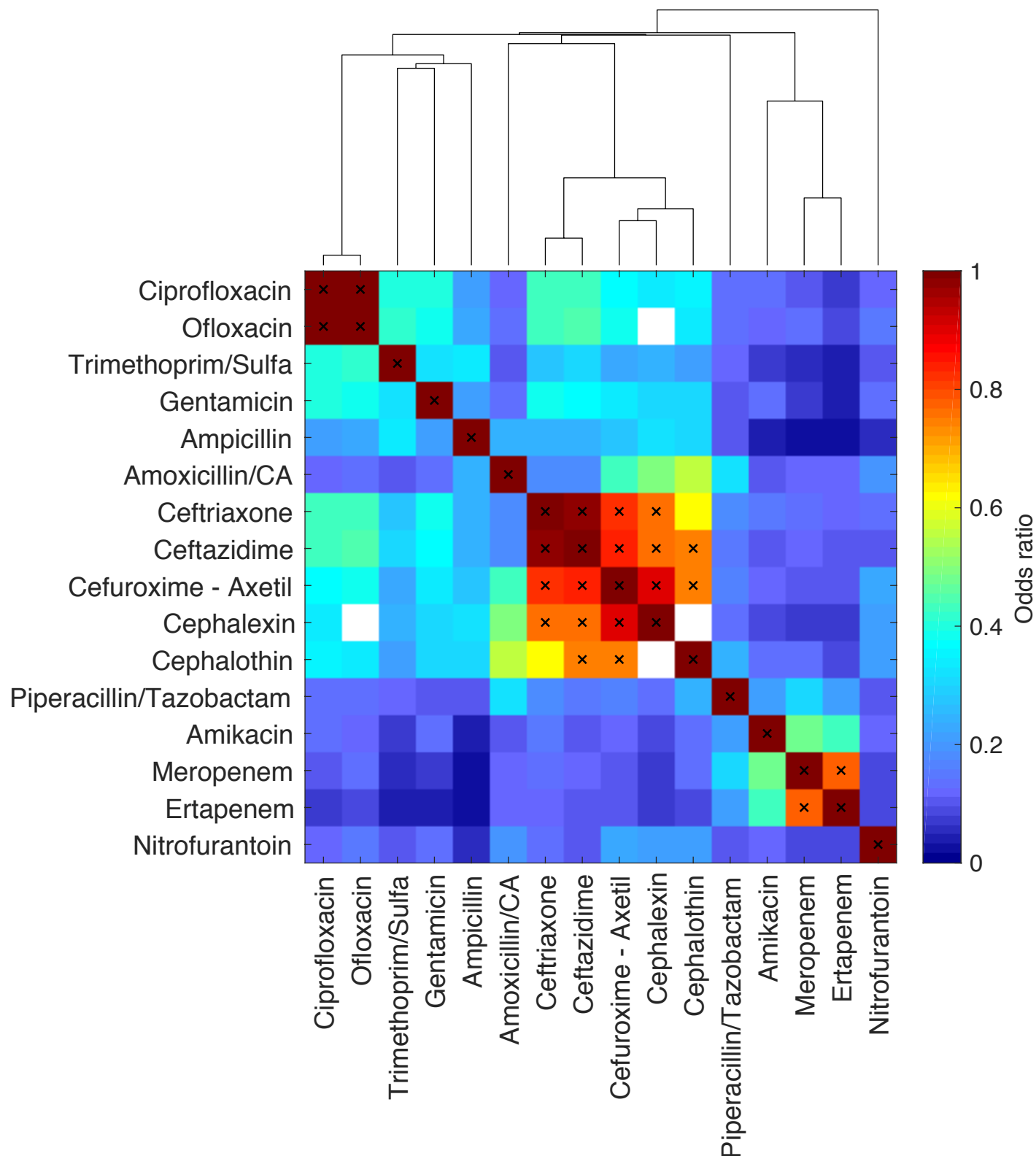
Supplementary Figure 6: Odds ratios of resistance to different antibiotics for past purchases of different drugs across a range of purchase-to-sample time intervals.
(a) Adjusted for all purchases



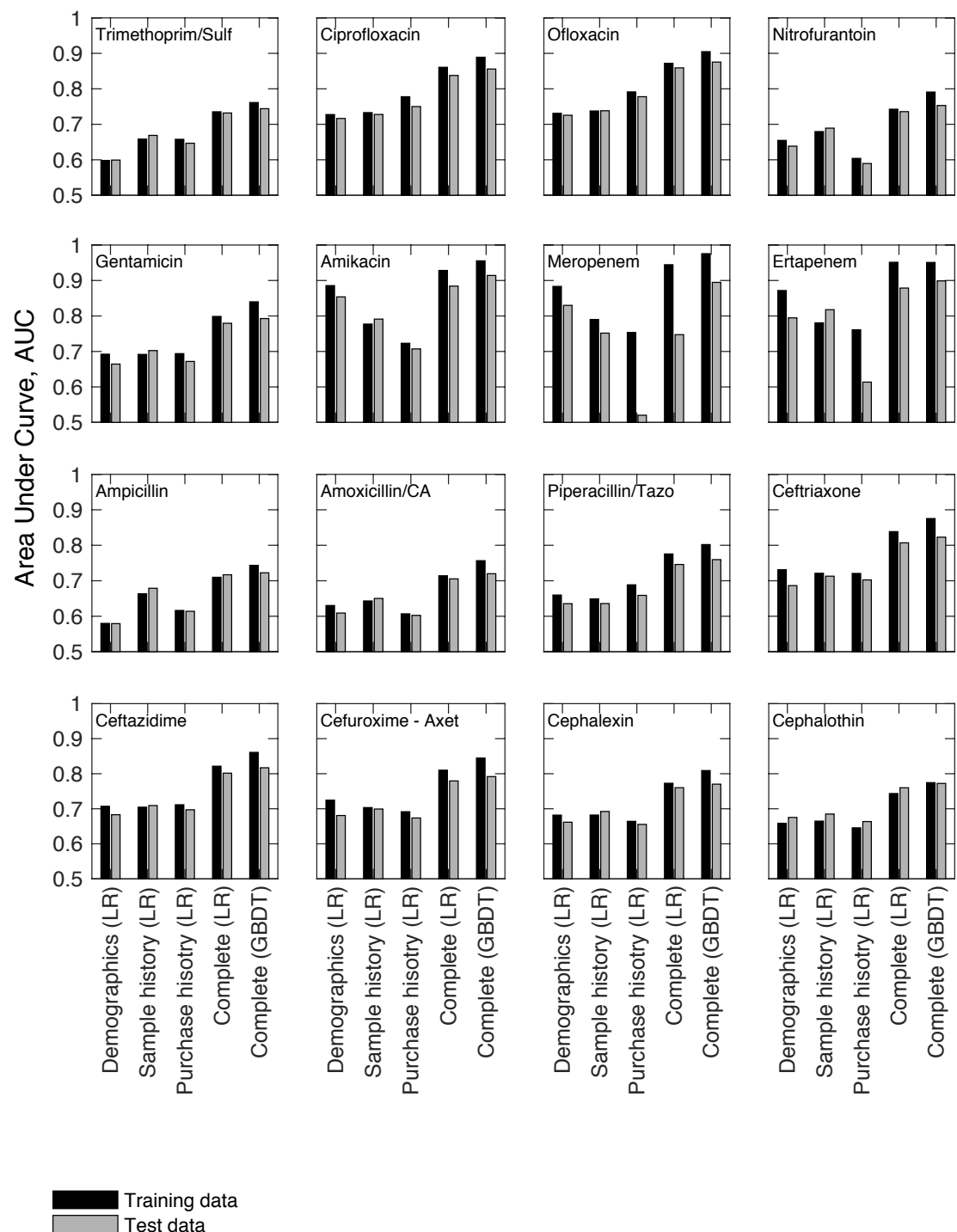
Supplementary Figure 6: Odds ratios of resistance to different antibiotics for past purchases of different drugs across a range of purchase-to-sample time intervals.
(b) Adjusted for all purchases and demographics



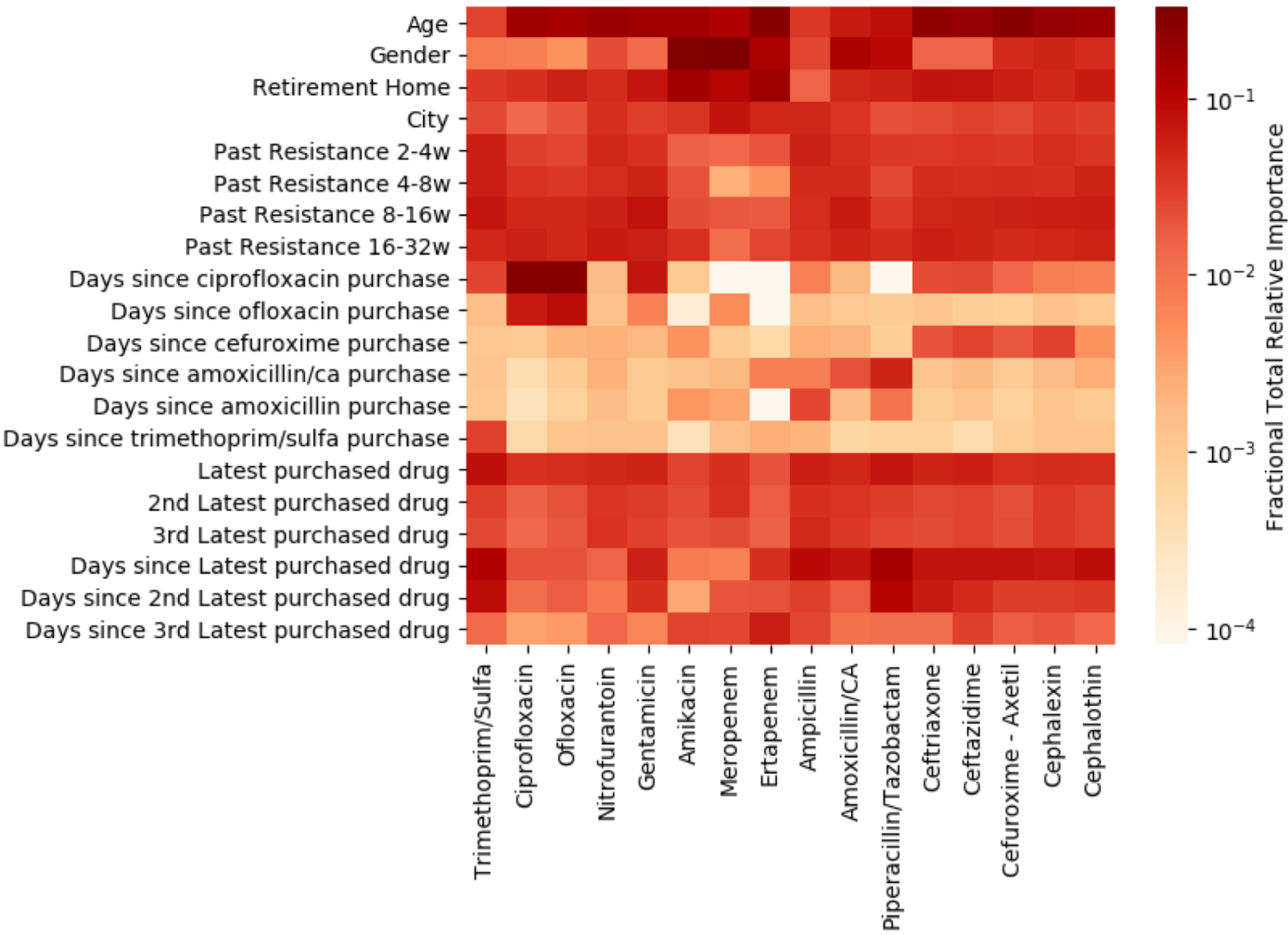
Supplementary Figure 6: Odds ratios of resistance to different antibiotics for past purchases of different drugs across a range of purchase-to-sample time intervals.
(c) Adjusted for all purchases and cross-resistance



Supplementary Figure 7: Correlations among resistances to different antibiotics.



Supplementary Figure 8: Model performance on test and training data.



Supplementary Figure 9: Feature importance in Gradient Boosting Decision Trees (GBDT).

Supplementary Figures

Supplementary Figure 1: Availability of resistance measurements over time. For each of the 16 antibiotics, the fraction of cultures for which resistance was measured, overall (black) and for each of the three most common species (colors) is plotted across the 10-year sampling period. Availability of measurements determined the time range used for model Training (green horizontal bars) and Testing (red bars) of each resistance. Time periods during which measurements of resistance to a given antibiotic were scarce were removed from analysis (gray bars).

Supplementary Figure 2: Frequency of resistance over time. Frequencies of resistance of each of the three common species (colored lines) and the overall sample (black lines) over the 10 year sampling period. Empty time intervals correspond to periods during which resistance was not frequently measured (matching the gray horizontal bars of **Supplementary Fig. 1**). The apparent surge in piperacillin/tazobactam resistance in 2009 is attributed to a switch in VITEK2 reference and does not represent a change in resistance frequency⁴³.

Supplementary Figure 3: Distribution of UTI demographics by city. Number (bottom) and fraction (top) of UTI samples by 5 demographic categories (legend), for the 20 cities with most samples. ‘Others’ bar for samples in all other cities.

Supplementary Figure 4: Odds of resistance as a function of age for different demographic groups. Odds of resistance to each of the 16 antibiotics, normalized to the overall population odds (η/η_0), in each of 10 age bins (0,10,...,100 years) for five non-overlapping demographic groups: men not residing in retirement homes (blue), men residing in retirement homes (dotted blue), women not pregnant and not residing in retirement homes (magenta), women in retirement homes (magenta dotted), and pregnant women (red).

Supplementary Figure 5: Odds ratio of resistance across same-patient sample pairs. Odds ratio for resistance to each of the 16 antibiotics, given previous resistant versus sensitive urine sample from the same patient, shown as a function of the time between the two samples ($OR_{pairs} = (N_{R \rightarrow R}/N_{R \rightarrow S}) / (N_{S \rightarrow R}/N_{S \rightarrow S})$, Methods). Odds ratios are fitted with $OR_{pairs} = C_0 e^{t/\tau_0} + C_m e^{t/\tau_m} + C_\infty$, representing short and long memory time scales and a timeless constant (red, yellow and green respectively). The magnitudes of these terms are shown as stacked red, yellow and green bars and the half life (τ_m) of

long time scale indicated (yellow arrow). The data of antibiotics amikacin, meropenem, ertapenem and piperacillin/tazobactam were too noisy to be reliably fitted.

Supplementary Figure 6: Odds ratios of resistance to each of the antibiotics for past purchases of different drugs across a range of purchase-to-sample time intervals. Adjusted odds ratio of resistance for purchase of antibiotics, $\exp(C_{k,i,j}^{ATC})$ for (a) model with purchases only (Methods: Logistic regression “Purchase history”), (b) model with purchases and demographics (Methods: Logistic regression “Purchase history adjusted for demographics”), (c) model disentangling indirect effects by accounting for linkage between resistances (Methods: Logistic regression “Purchase history adjusted for cross-resistance”). Same graphical scheme as Fig. 4a,b. Gray asterisks indicate statistical significance.

Supplementary Figure 7: Correlations among resistances to different antibiotics. Correlation among resistance measurements for each pair of antibiotics across all samples for which both resistances were measured (colormap; empty white squares correspond to antibiotic pairs measured in non-overlapping time periods, **Supplementary Fig. 1**). Correlations values vary between 0.99 (Ciprofloxacin-Ofloxacin) and 0.03 (Ertapenem-Ampicillin). Even the weaker correlations are highly significant (all correlations have $P < 10^{-20}$). Pairs of antibiotics with correlation values higher than 0.7 are treated as “analogous”, and their correlations are excluded from analysis of indirect effects of purchases on resistance (Methods: Logistic regression “Purchase history adjusted for cross-resistance”).

Supplementary Figure 8: Model performance on test and training data. Area Under Curve (AUC) for Receiver Operator Characteristic for prediction of resistance based on demographics, sample history and purchase history, individually and in a complete model combining all feature sets. Each feature set was modelled using Logistic Regression (LR), and the complete model was modelled by both LR and Gradient Boosting Decision Trees (GBDT). To identify overfitting, model performance on the testing dataset (grey) was contrasted with model performance on the training dataset (black; **Supplementary Fig. 1** for definition of training and test time periods). For most antibiotics both LR and GBDT models had similar predictive power when tested on the training and the testing datasets, suggesting no significant overfitting. We could however detect overfitting for the LR model of meropenem, especially in the purchase history model, corresponding to the small number of purchases of this drug (**Supplementary Table 2**). This overfitting was substantially reduced in the GBDT model, where the hyperparameter tuning determined a relatively low optimal decision tree depth limit (3 compared to 6-7 for most other antibiotics), lowering the variance error, and reducing overfitting.

Supplementary Figure 9: Feature importance in Gradient Boosting Decision Trees (GBDT). The relative importance of each feature on the prediction of the GBDT model of each antibiotic is shown (color map)⁴⁴. Importance is defined, for each feature, as its fractional estimated total error improvement, which is the sum of error difference over all instances of the feature in all trees in an ensemble⁴⁴. For each antibiotic, total feature error improvements are normalized to one. The features shown are those of highest influence for each of the main features sets: demographics, sample history and drug purchase history.

Supplementary Table 1: Regression coefficients for association of resistance with demographics. Logistic regression of resistance to each of the 16 antibiotics with respect to each of the demographic features alone (unadjusted) and when accounted for together (adjusted). For definitions of listed demographic terms see Methods.

Age_1	-1.51 [-1.60,-1.42]***	-1.09 [-1.18,-1.00]***	2.31 [2.11, 2.51]***	2.18 [1.97, 2.38]***
Age_2	2.74 [2.65, 2.83]***	2.05 [1.96, 2.15]***	1.20 [1.03, 1.37]***	0.63 [0.46, 0.81]***
Ret.Home	1.15 [1.12, 1.18]***	0.66 [0.63, 0.69]***	1.87 [1.84, 1.90]***	1.01 [0.98, 1.04]***
Gender	0.30 [0.28, 0.31]***	0.06 [0.04, 0.08]***	1.05 [1.03, 1.06]***	0.61 [0.59, 0.63]***
Pregnancy	-0.71 [-0.74,-0.68]***	-0.48 [-0.50,-0.45]***	-1.54 [-1.59,-1.50]***	-0.77 [-0.82,-0.72]***
Date_1	0.03 [0.03, 0.04]***	0.03 [0.02, 0.04]***	0.09 [0.08, 0.10]***	0.07 [0.06, 0.08]***
Date_2	-0.00 [-0.01,-0.00]***	-0.00 [-0.01,-0.00]***	-0.01 [-0.01,-0.01]***	-0.01 [-0.01,-0.01]***
Season_1	-0.03 [-0.04,-0.02]***	-0.02 [-0.03,-0.01]**	-0.03 [-0.04,-0.02]***	-0.02 [-0.03,-0.01]**
Season_2	0.01 [-0.00, 0.01]	0.01 [0.00, 0.02]	-0.01 [-0.02,-0.00]	-0.01 [-0.02, 0.00]
Tel-Aviv-Reference	0	0	0	0
City_Haifa	0.27 [0.24, 0.30]***	0.11 [0.08, 0.14]***	0.32 [0.29, 0.35]***	-0.04 [-0.08,-0.00]
City_Rishon le Zion	-0.12 [-0.16,-0.09]***	-0.03 [-0.07,-0.00]	-0.33 [-0.37,-0.30]***	-0.14 [-0.18,-0.10]***
City_Holon	0.03 [0.00, 0.06]	0.09 [0.06, 0.12]***	-0.07 [-0.11,-0.04]**	0.04 [0.00, 0.08]
City_Jerusalem	-0.03 [-0.06, 0.00]	-0.02 [-0.05, 0.01]	-0.22 [-0.26,-0.17]***	-0.22 [-0.26,-0.17]***
City_Ramat-Hasharon	-0.13 [-0.16,-0.10]***	-0.10 [-0.14,-0.07]***	-0.25 [-0.30,-0.21]***	-0.23 [-0.27,-0.18]***
City_Petah Tiqva	-0.05 [-0.08,-0.01]*	-0.00 [-0.04, 0.03]	-0.20 [-0.24,-0.15]***	-0.13 [-0.17,-0.08]***
City_Bat-Yam	0.39 [0.36, 0.42]***	0.24 [0.21, 0.27]***	0.61 [0.57, 0.64]***	0.30 [0.26, 0.33]***
City_Ashdod	0.03 [-0.01, 0.07]	0.02 [-0.02, 0.05]	-0.03 [-0.08, 0.01]	-0.11 [-0.15,-0.06]**
City_Bney Brak	-0.05 [-0.08,-0.01]	-0.01 [-0.04, 0.03]	-0.10 [-0.15,-0.05]**	-0.04 [-0.09, 0.01]
City_Beer-Sheva	-0.04 [-0.08,-0.00]	0.00 [-0.04, 0.04]	-0.08 [-0.13,-0.04]*	0.01 [-0.04, 0.06]
City_Netanya	0.11 [0.07, 0.14]***	0.03 [-0.01, 0.07]	0.22 [0.18, 0.26]***	0.02 [-0.03, 0.06]
City_Ramat Gan	-0.07 [-0.11,-0.03]*	-0.08 [-0.12,-0.04]**	-0.01 [-0.05, 0.04]	-0.04 [-0.09, 0.01]
City_Kiryat Motzkin	0.05 [0.00, 0.09]	0.05 [0.01, 0.09]	-0.00 [-0.06, 0.05]	-0.03 [-0.08, 0.03]
City_Raanana	0.10 [0.05, 0.14]**	0.04 [-0.01, 0.08]	0.06 [0.01, 0.11]	-0.12 [-0.18,-0.06]**
City_Kfar Saba	0.05 [0.00, 0.10]	0.01 [-0.04, 0.06]	-0.01 [-0.07, 0.05]	-0.14 [-0.20,-0.08]**
City_Modiin	-0.21 [-0.26,-0.15]***	-0.05 [-0.10, 0.01]	-0.65 [-0.73,-0.57]***	-0.26 [-0.34,-0.18]***
City_Ashkelon	0.07 [0.02, 0.12]*	0.08 [0.03, 0.13]*	0.20 [0.14, 0.26]***	0.17 [0.11, 0.23]***
City_Givaataim	-0.10 [-0.16,-0.05]*	-0.04 [-0.10, 0.01]	-0.27 [-0.33,-0.20]***	-0.15 [-0.22,-0.08]**
City_Kiryat Ono	-0.04 [-0.09, 0.01]	-0.07 [-0.12,-0.01]	0.04 [-0.02, 0.10]	-0.04 [-0.10, 0.03]
City_Others	-0.06 [-0.08,-0.03]***	0.03 [0.01, 0.05]	-0.21 [-0.23,-0.18]***	-0.05 [-0.07,-0.02]*

	Ofloxacin		Nitrofurantoin	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Age_1	2.30 [1.96, 2.64]***	2.34 [1.99, 2.69]***	-4.20 [-4.32,-4.08]***	-3.73 [-3.85,-3.61]***
Age_2	1.36 [1.08, 1.64]***	0.60 [0.30, 0.90]**	5.61 [5.49, 5.73]***	4.66 [4.53, 4.79]***
Ret.Home	2.04 [2.00, 2.08]***	1.14 [1.09, 1.18]***	1.65 [1.62, 1.68]***	0.99 [0.95, 1.02]***
Gender	1.02 [0.99, 1.04]***	0.59 [0.56, 0.62]***	0.87 [0.85, 0.89]***	0.65 [0.63, 0.67]***
Pregnancy	-1.61 [-1.68,-1.54]***	-0.76 [-0.84,-0.69]***	-0.25 [-0.28,-0.21]***	0.11 [0.08, 0.15]***
Date_1	0.15 [0.13, 0.18]***	0.16 [0.14, 0.19]***	0.02 [0.01, 0.03]**	0.03 [0.02, 0.04]**
Date_2	-0.02 [-0.02,-0.01]***	-0.02 [-0.02,-0.01]***	-0.00 [-0.00, 0.00]	-0.00 [-0.00, 0.00]
Season_1	-0.04 [-0.05,-0.02]***	-0.05 [-0.07,-0.04]***	0.07 [0.06, 0.08]***	0.07 [0.05, 0.08]***
Season_2	-0.01 [-0.03, 0.00]	-0.02 [-0.04,-0.01]*	0.07 [0.05, 0.08]***	0.07 [0.06, 0.08]***
Tel-Aviv-Reference	0	0	0	0
City_Haifa	0.34 [0.30, 0.39]***	-0.01 [-0.07, 0.04]	0.45 [0.41, 0.49]***	0.19 [0.15, 0.24]***
City_Rishon le Zion	-0.37 [-0.43,-0.31]***	-0.16 [-0.21,-0.10]***	-0.04 [-0.09, 0.00]	0.08 [0.04, 0.13]*
City_Holon	-0.09 [-0.15,-0.04]*	0.05 [-0.00, 0.11]	0.05 [0.01, 0.10]	0.16 [0.11, 0.20]***
City_Jerusalem	-0.19 [-0.25,-0.13]***	-0.19 [-0.25,-0.13]***	0.01 [-0.04, 0.06]	-0.01 [-0.06, 0.04]
City_Ramat-Hasharon	-0.24 [-0.30,-0.18]***	-0.24 [-0.31,-0.17]***	-0.07 [-0.12,-0.02]*	-0.07 [-0.12,-0.02]*
City_Petah Tiqva	-0.19 [-0.26,-0.13]***	-0.16 [-0.22,-0.09]**	0.03 [-0.02, 0.08]	0.04 [-0.01, 0.09]
City_Bat-Yam	0.64 [0.59, 0.69]***	0.33 [0.28, 0.39]***	0.63 [0.59, 0.68]***	0.42 [0.37, 0.46]***
City_Ashdod	0.01 [-0.06, 0.07]	-0.07 [-0.13,-0.00]	0.21 [0.16, 0.26]***	0.17 [0.11, 0.22]***
City_Bney Brak	0.01 [-0.06, 0.07]	-0.04 [-0.11, 0.03]	0.19 [0.14, 0.24]***	0.12 [0.07, 0.17]**
City_Beer-Sheva	-0.01 [-0.08, 0.05]	0.05 [-0.01, 0.12]	0.07 [0.01, 0.12]	0.12 [0.07, 0.18]**
City_Netanya	0.31 [0.25, 0.37]***	0.10 [0.03, 0.16]*	0.43 [0.38, 0.48]***	0.29 [0.24, 0.34]***
City_Ramat Gan	-0.09 [-0.15,-0.02]*	-0.09 [-0.16,-0.02]*	-0.04 [-0.10, 0.02]	-0.02 [-0.07, 0.04]
City_Kiryat Motzkin	-0.01 [-0.09, 0.06]	-0.02 [-0.10, 0.06]	0.18 [0.12, 0.24]***	0.19 [0.13, 0.25]***
City_Raanana	0.06 [-0.02, 0.14]	-0.11 [-0.19,-0.03]*	0.19 [0.13, 0.26]***	0.09 [0.02, 0.15]*
City_Kfar Saba	-0.08 [-0.16, 0.00]	-0.21 [-0.30,-0.12]**	0.16 [0.10, 0.23]**	0.09 [0.02, 0.16]*
City_Modiin	-0.61 [-0.73,-0.50]***	-0.18 [-0.31,-0.06]*	-0.15 [-0.23,-0.07]**	0.04 [-0.04, 0.12]
City_Ashkelon	0.14 [0.06, 0.23]*	0.09 [-0.01, 0.18]	0.31 [0.25, 0.38]***	0.28 [0.21, 0.35]***
City_Givaataim	-0.32 [-0.41,-0.22]***	-0.15 [-0.25,-0.06]*	-0.20 [-0.28,-0.11]**	-0.09 [-0.17,-0.00]
City_Kiryat Ono	0.00 [-0.09, 0.09]	-0.08 [-0.18, 0.01]	0.09 [0.01, 0.16]	0.03 [-0.05, 0.10]
City_Others	-0.26 [-0.30,-0.22]***	-0.07 [-0.11,-0.04]*	0.00 [-0.03, 0.03]	0.11 [0.07, 0.14]***

Age_1	-0.44 [-0.60, -0.27]***	0.23 [0.06, 0.40]*	1.39 [0.46, 2.31]*	1.75 [0.83, 2.67]*
Age_2	2.76 [2.60, 2.91]***	1.39 [1.22, 1.55]***	2.97 [2.20, 3.73]***	0.78 [-0.00, 1.55]
Ret.Home	1.91 [1.88, 1.94]***	1.20 [1.16, 1.23]***	2.89 [2.81, 2.97]***	1.61 [1.51, 1.70]***
Gender	1.05 [1.03, 1.07]***	0.69 [0.66, 0.71]***	2.46 [2.38, 2.53]***	1.87 [1.79, 1.95]***
Pregnancy	-0.94 [-0.99, -0.88]***	-0.50 [-0.56, -0.44]***	-1.40 [-1.82, -0.98]***	-0.59 [-1.02, -0.16]*
Date_1	0.10 [0.08, 0.11]***	0.11 [0.09, 0.12]***	-0.05 [-0.10, 0.00]	-0.04 [-0.10, 0.01]
Date_2	-0.01 [-0.01, -0.00]***	-0.01 [-0.01, -0.01]***	-0.01 [-0.01, -0.00]	-0.01 [-0.01, -0.00]
Season_1	-0.00 [-0.02, 0.01]	-0.01 [-0.02, 0.00]	-0.02 [-0.08, 0.03]	0.01 [-0.04, 0.07]
Season_2	0.00 [-0.01, 0.01]	0.01 [-0.00, 0.02]	-0.05 [-0.11, -0.00]	-0.04 [-0.09, 0.01]
Tel-Aviv-Reference	0	0	0	0
City_Haifa	0.40 [0.36, 0.45]***	0.04 [-0.01, 0.09]	0.77 [0.60, 0.95]***	0.09 [-0.09, 0.27]
City_Rishon le Zion	-0.16 [-0.21, -0.11]***	0.01 [-0.04, 0.06]	-0.99 [-1.29, -0.68]***	-0.60 [-0.91, -0.30]**
City_Holon	-0.01 [-0.06, 0.04]	0.11 [0.05, 0.16]**	-0.42 [-0.67, -0.17]*	-0.12 [-0.38, 0.13]
City_Jerusalem	-0.12 [-0.18, -0.07]**	-0.14 [-0.19, -0.08]**	-0.11 [-0.36, 0.13]	-0.20 [-0.45, 0.05]
City_Ramat-Hasharon	-0.28 [-0.34, -0.22]***	-0.23 [-0.29, -0.17]***	-0.96 [-1.30, -0.62]***	-0.87 [-1.21, -0.53]***
City_Petah Tiqva	-0.22 [-0.27, -0.16]***	-0.15 [-0.21, -0.09]***	-0.08 [-0.33, 0.16]	-0.05 [-0.30, 0.20]
City_Bat-Yam	0.76 [0.71, 0.81]***	0.42 [0.37, 0.47]***	1.65 [1.49, 1.81]***	0.92 [0.75, 1.09]***
City_Ashdod	0.08 [0.02, 0.14]*	0.02 [-0.04, 0.08]	0.21 [-0.04, 0.45]	0.03 [-0.22, 0.28]
City_Bney Brak	-0.09 [-0.15, -0.03]*	-0.08 [-0.15, -0.01]	0.36 [0.13, 0.59]*	0.23 [-0.00, 0.47]
City_Beer-Sheva	-0.04 [-0.10, 0.03]	0.06 [-0.01, 0.12]	0.23 [-0.02, 0.47]	0.43 [0.19, 0.68]*
City_Netanya	0.33 [0.28, 0.39]***	0.13 [0.07, 0.19]**	0.54 [0.32, 0.76]***	0.15 [-0.07, 0.38]
City_Ramat Gan	-0.04 [-0.11, 0.02]	-0.05 [-0.11, 0.02]	-0.06 [-0.34, 0.22]	-0.00 [-0.29, 0.28]
City_Kiryat Motzkin	0.05 [-0.02, 0.12]	0.02 [-0.05, 0.09]	-0.00 [-0.31, 0.30]	-0.10 [-0.40, 0.21]
City_Raanana	0.25 [0.18, 0.32]***	0.05 [-0.02, 0.12]	0.54 [0.28, 0.80]**	0.06 [-0.20, 0.33]
City_Kfar Saba	0.13 [0.05, 0.21]*	-0.01 [-0.09, 0.06]	0.16 [-0.17, 0.48]	-0.17 [-0.50, 0.16]
City_Modiin	-0.39 [-0.49, -0.30]***	-0.08 [-0.18, 0.01]	-2.01 [-2.91, -1.10]**	-1.23 [-2.14, -0.33]*
City_Ashkelon	0.27 [0.20, 0.35]***	0.23 [0.15, 0.31]***	-0.34 [-0.75, 0.07]	-0.52 [-0.94, -0.10]
City_Givaataim	-0.16 [-0.26, -0.07]*	-0.05 [-0.14, 0.05]	-1.55 [-2.32, -0.79]**	-1.24 [-2.01, -0.47]*
City_Kiryat Ono	-0.00 [-0.09, 0.09]	-0.11 [-0.20, -0.02]	-0.27 [-0.69, 0.15]	-0.66 [-1.08, -0.23]*
City_Others	-0.12 [-0.16, -0.09]***	0.01 [-0.03, 0.05]	-0.27 [-0.43, -0.12]*	0.01 [-0.15, 0.17]

	Meropenem		Ertapenem	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Age_1	2.43 [0.30, 4.56]	2.35 [0.27, 4.43]	-0.04 [-1.74, 1.67]	1.64 [-0.12, 3.40]
Age_2	1.83 [0.08, 3.58]	0.11 [-1.63, 1.85]	4.27 [2.86, 5.69]***	1.09 [-0.39, 2.58]
Ret.Home	2.63 [2.44, 2.82]***	1.47 [1.24, 1.70]***	2.92 [2.76, 3.07]***	1.65 [1.46, 1.84]***
Gender	2.63 [2.46, 2.81]***	2.10 [1.91, 2.28]***	1.95 [1.80, 2.10]***	1.33 [1.18, 1.49]***
Pregnancy	-0.94 [-1.77, -0.11]	-0.26 [-1.11, 0.58]	-1.39 [-2.11, -0.68]**	-0.49 [-1.22, 0.24]
Date_1	-0.90 [-2.02, 0.21]	-0.80 [-1.92, 0.32]	-0.10 [-0.28, 0.09]	-0.09 [-0.28, 0.09]
Date_2	0.05 [-0.02, 0.12]	0.04 [-0.03, 0.11]	0.00 [-0.01, 0.02]	0.00 [-0.01, 0.02]
Season_1	-0.06 [-0.18, 0.07]	-0.03 [-0.16, 0.10]	-0.09 [-0.20, 0.02]	-0.06 [-0.16, 0.05]
Season_2	-0.16 [-0.28, -0.04]*	-0.14 [-0.26, -0.01]	-0.14 [-0.24, -0.03]*	-0.12 [-0.22, -0.01]
Tel-Aviv-Reference	0	0	0	0
City_Haifa	1.30 [0.85, 1.75]***	0.54 [0.08, 0.99]	1.42 [1.10, 1.75]***	0.78 [0.45, 1.11]**
City_Rishon le Zion	-0.29 [-0.94, 0.37]	-0.03 [-0.68, 0.63]	-0.74 [-1.31, -0.17]*	-0.34 [-0.92, 0.23]
City_Holon	0.16 [-0.43, 0.75]	0.31 [-0.28, 0.90]	-0.45 [-0.97, 0.08]	-0.15 [-0.67, 0.38]
City_Jerusalem	0.14 [-0.49, 0.77]	0.00 [-0.63, 0.63]	0.23 [-0.22, 0.67]	0.18 [-0.27, 0.63]
City_Ramat-Hasharon	-0.42 [-1.19, 0.35]	-0.35 [-1.12, 0.42]	-0.66 [-1.28, -0.04]	-0.63 [-1.25, -0.01]
City_Petah Tiqva	0.30 [-0.30, 0.91]	0.32 [-0.29, 0.93]	-0.42 [-0.99, 0.15]	-0.36 [-0.93, 0.21]
City_Bat-Yam	1.41 [0.94, 1.89]***	0.68 [0.19, 1.17]*	1.17 [0.80, 1.53]***	0.46 [0.09, 0.83]
City_Ashdod	0.49 [-0.14, 1.11]	0.28 [-0.35, 0.91]	-0.19 [-0.78, 0.39]	-0.31 [-0.90, 0.28]
City_Bney Brak	1.12 [0.59, 1.64]**	1.10 [0.57, 1.63]**	1.10 [0.71, 1.49]***	1.02 [0.63, 1.41]***
City_Beer-Sheva	-0.24 [-1.04, 0.57]	-0.00 [-0.81, 0.80]	0.14 [-0.38, 0.66]	0.40 [-0.12, 0.92]
City_Netanya	1.52 [1.04, 1.99]***	1.07 [0.59, 1.55]**	0.81 [0.40, 1.23]**	0.44 [0.02, 0.86]
City_Ramat Gan	1.00 [0.43, 1.57]*	0.96 [0.39, 1.54]*	0.39 [-0.11, 0.89]	0.45 [-0.05, 0.95]
City_Kiryat Motzkin	0.13 [-0.67, 0.93]	0.03 [-0.77, 0.84]	0.08 [-0.52, 0.68]	0.03 [-0.58, 0.63]
City_Raanana	1.39 [0.84, 1.94]***	0.91 [0.35, 1.47]*	0.74 [0.24, 1.24]*	0.32 [-0.18, 0.83]
City_Kfar Saba	0.69 [-0.05, 1.42]	0.42 [-0.33, 1.16]	0.35 [-0.27, 0.96]	0.03 [-0.59, 0.65]
City_Modiin	-0.59 [-1.81, 0.63]	-0.05 [-1.26, 1.17]	-0.83 [-1.86, 0.21]	-0.12 [-1.16, 0.92]
City_Ashkelon	0.15 [-0.75, 1.05]	-0.01 [-0.91, 0.90]	-1.07 [-2.26, 0.12]	-1.22 [-2.41, -0.03]
City_Givaataim	-1.41 [-3.44, 0.63]	-1.19 [-3.23, 0.85]	-1.34 [-2.78, 0.10]	-1.03 [-2.47, 0.42]
City_Kiryat Ono	-0.70 [-2.16, 0.76]	-1.05 [-2.52, 0.42]	0.05 [-0.71, 0.81]	-0.29 [-1.05, 0.48]
City_Others	0.36 [-0.05, 0.77]	0.52 [0.11, 0.94]	-0.22 [-0.54, 0.10]	0.07 [-0.25, 0.39]

Age_1	-0.85 [-0.97,-0.73]***	-0.65 [-0.78,-0.53]***	-1.47 [-1.61,-1.33]***	-1.01 [-1.15,-0.86]***
Age_2	1.90 [1.78, 2.02]***	1.47 [1.34, 1.60]***	2.79 [2.65, 2.93]***	1.78 [1.63, 1.93]***
Ret.Home	1.05 [1.00, 1.11]***	0.64 [0.58, 0.70]***	1.41 [1.37, 1.44]***	0.88 [0.84, 0.92]***
Gender	0.70 [0.68, 0.73]***	0.55 [0.52, 0.57]***	1.08 [1.06, 1.10]***	0.90 [0.87, 0.92]***
Pregnancy	-0.14 [-0.17,-0.12]***	0.02 [-0.02, 0.05]	-0.19 [-0.23,-0.15]***	0.03 [-0.01, 0.07]
Date_1	-0.21 [-0.29,-0.12]***	-0.23 [-0.32,-0.15]***	-0.05 [-0.07,-0.04]***	-0.05 [-0.06,-0.04]***
Date_2	0.01 [0.01, 0.02]**	0.01 [0.01, 0.02]**	0.01 [0.00, 0.01]***	0.00 [0.00, 0.01]***
Season_1	-0.01 [-0.02,-0.00]	-0.01 [-0.03,-0.00]*	-0.03 [-0.04,-0.01]**	-0.03 [-0.04,-0.01]*
Season_2	0.01 [-0.00, 0.02]	0.02 [0.00, 0.03]*	-0.01 [-0.02, 0.01]	-0.00 [-0.01, 0.01]
Tel-Aviv-Reference	0	0	0	0
City_Haifa	0.20 [0.16, 0.24]***	0.06 [0.01, 0.10]*	0.37 [0.32, 0.41]***	0.13 [0.09, 0.18]***
City_Rishon le Zion	-0.03 [-0.07, 0.01]	0.02 [-0.02, 0.06]	-0.16 [-0.21,-0.11]***	-0.07 [-0.12,-0.02]*
City_Holon	0.13 [-0.09, 0.17]***	0.15 [0.10, 0.19]***	-0.03 [-0.08, 0.02]	0.02 [-0.03, 0.07]
City_Jerusalem	0.07 [0.03, 0.11]*	0.08 [0.03, 0.12]*	0.06 [0.01, 0.12]	0.05 [-0.00, 0.10]
City_Ramat-Hasharon	-0.07 [-0.12,-0.03]*	-0.04 [-0.09,-0.00]	-0.12 [-0.18,-0.07]**	-0.10 [-0.16,-0.05]*
City_Petah Tiqva	-0.06 [-0.10,-0.01]*	-0.02 [-0.07, 0.02]	-0.04 [-0.10, 0.01]	-0.02 [-0.07, 0.04]
City_Bat-Yam	0.35 [0.30, 0.40]***	0.22 [0.17, 0.27]***	0.40 [0.35, 0.45]***	0.16 [0.11, 0.22]***
City_Ashdod	0.03 [-0.02, 0.08]	-0.02 [-0.07, 0.03]	0.05 [-0.01, 0.11]	-0.01 [-0.07, 0.05]
City_Bney Brak	0.19 [0.14, 0.24]***	0.23 [0.18, 0.28]***	0.15 [0.10, 0.21]***	0.14 [0.08, 0.20]**
City_Beer-Sheva	-0.01 [-0.06, 0.03]	0.02 [-0.03, 0.07]	0.01 [-0.05, 0.07]	0.05 [-0.01, 0.11]
City_Netanya	0.21 [0.16, 0.26]***	0.15 [0.10, 0.20]***	0.18 [0.12, 0.23]***	0.04 [-0.02, 0.10]
City_Ramat Gan	0.08 [0.02, 0.13]*	0.05 [-0.00, 0.10]	-0.08 [-0.14,-0.02]	-0.08 [-0.15,-0.02]
City_Kiryat Motzkin	0.06 [0.01, 0.12]	0.06 [0.00, 0.11]	0.08 [0.01, 0.14]	0.06 [-0.01, 0.13]
City_Raanana	0.14 [0.08, 0.20]**	0.09 [0.03, 0.15]*	0.17 [0.10, 0.24]**	0.06 [-0.01, 0.13]
City_Kfar Saba	0.02 [-0.04, 0.09]	-0.02 [-0.08, 0.05]	0.08 [0.00, 0.15]	-0.00 [-0.08, 0.07]
City_Modiin	-0.16 [-0.22,-0.10]***	-0.06 [-0.12, 0.01]	-0.31 [-0.40,-0.22]***	-0.14 [-0.23,-0.05]*
City_Ashkelon	0.09 [0.03, 0.16]*	0.07 [0.01, 0.14]	-0.01 [-0.09, 0.07]	-0.06 [-0.14, 0.02]
City_Givaataim	0.05 [-0.02, 0.12]	0.07 [-0.00, 0.14]	-0.08 [-0.17, 0.01]	-0.01 [-0.10, 0.08]
City_Kiryat Ono	0.06 [-0.01, 0.13]	0.04 [-0.03, 0.11]	-0.05 [-0.14, 0.03]	-0.12 [-0.21,-0.03]*
City_Others	0.06 [0.03, 0.09]**	0.10 [0.07, 0.13]***	-0.04 [-0.07,-0.00]	0.04 [0.00, 0.07]

	Piperacillin/Ta		Ceftriaxone	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Age_1	-1.68 [-2.16,-1.21]***	-1.20 [-1.68,-0.72]***	-0.48 [-0.67,-0.30]***	0.43 [0.24, 0.62]**
Age_2	3.27 [2.81, 3.73]***	2.20 [1.72, 2.68]***	3.34 [3.17, 3.50]***	1.66 [1.48, 1.84]***
Ret.Home	1.46 [1.36, 1.56]***	0.81 [0.70, 0.93]***	2.39 [2.36, 2.42]***	1.59 [1.56, 1.63]***
Gender	1.16 [1.09, 1.22]***	0.90 [0.83, 0.97]***	1.17 [1.15, 1.19]***	0.75 [0.73, 0.78]***
Pregnancy	-0.43 [-0.59,-0.28]***	-0.16 [-0.32, 0.00]	-0.80 [-0.86,-0.75]***	-0.21 [-0.27,-0.15]***
Date_1	-0.42 [-0.92, 0.09]	-0.28 [-0.80, 0.24]	0.06 [0.04, 0.08]***	0.07 [0.05, 0.09]***
Date_2	0.03 [-0.00, 0.06]	0.02 [-0.01, 0.05]	0.00 [-0.00, 0.00]	-0.00 [-0.00, 0.00]
Season_1	0.03 [-0.01, 0.08]	0.03 [-0.02, 0.07]	0.00 [-0.01, 0.02]	-0.02 [-0.03,-0.00]
Season_2	0.05 [0.01, 0.09]	0.06 [0.02, 0.10]*	0.03 [0.02, 0.04]**	0.04 [0.03, 0.06]***
Tel-Aviv-Reference	0	0	0	0
City_Haifa	0.57 [0.43, 0.72]***	0.27 [0.12, 0.42]*	0.52 [0.47, 0.56]***	0.06 [0.01, 0.11]
City_Rishon le Zion	-0.11 [-0.29, 0.06]	-0.01 [-0.19, 0.16]	-0.29 [-0.35,-0.23]***	-0.08 [-0.14,-0.02]*
City_Holon	0.10 [-0.07, 0.26]	0.14 [-0.03, 0.31]	-0.06 [-0.12,-0.01]	0.08 [0.02, 0.13]*
City_Jerusalem	0.19 [0.01, 0.36]	0.17 [-0.01, 0.34]	0.04 [-0.02, 0.10]	0.03 [-0.03, 0.09]
City_Ramat-Hasharon	-0.06 [-0.25, 0.13]	-0.04 [-0.23, 0.16]	-0.36 [-0.42,-0.29]***	-0.32 [-0.39,-0.26]***
City_Petah Tiqva	0.08 [-0.11, 0.26]	0.11 [-0.08, 0.29]	-0.14 [-0.20,-0.08]**	-0.07 [-0.13,-0.01]
City_Bat-Yam	0.57 [0.40, 0.73]***	0.29 [0.12, 0.45]*	0.82 [0.77, 0.87]***	0.40 [0.35, 0.45]***
City_Ashdod	0.25 [0.06, 0.45]*	0.17 [-0.02, 0.36]	0.06 [-0.01, 0.12]	-0.03 [-0.09, 0.04]
City_Bney Brak	0.29 [0.10, 0.48]*	0.32 [0.13, 0.51]*	0.07 [0.00, 0.13]	0.06 [-0.01, 0.13]
City_Beer-Sheva	-0.06 [-0.27, 0.16]	0.02 [-0.19, 0.24]	-0.13 [-0.20,-0.06]**	-0.01 [-0.08, 0.06]
City_Netanya	0.56 [0.39, 0.73]***	0.40 [0.22, 0.57]**	0.29 [0.23, 0.35]***	0.02 [-0.04, 0.09]
City_Ramat Gan	0.03 [-0.19, 0.26]	0.00 [-0.22, 0.23]	-0.12 [-0.19,-0.05]*	-0.12 [-0.19,-0.05]*
City_Kiryat Motzkin	0.12 [-0.11, 0.35]	0.10 [-0.12, 0.33]	-0.04 [-0.12, 0.03]	-0.08 [-0.16,-0.00]
City_Raanana	0.30 [0.08, 0.53]*	0.17 [-0.06, 0.39]	0.27 [0.20, 0.34]***	0.01 [-0.07, 0.08]
City_Kfar Saba	0.01 [-0.27, 0.28]	-0.09 [-0.36, 0.19]	0.11 [0.03, 0.19]*	-0.10 [-0.19,-0.01]
City_Modiin	-0.13 [-0.41, 0.15]	0.06 [-0.22, 0.34]	-0.40 [-0.50,-0.30]***	-0.02 [-0.13, 0.08]
City_Ashkelon	0.14 [-0.11, 0.40]	0.10 [-0.16, 0.36]	0.22 [0.14, 0.30]***	0.13 [0.05, 0.22]*
City_Givaataim	-0.08 [-0.39, 0.23]	-0.01 [-0.32, 0.30]	-0.25 [-0.35,-0.15]***	-0.12 [-0.22,-0.02]
City_Kiryat Ono	0.01 [-0.30, 0.31]	-0.06 [-0.36, 0.25]	0.12 [0.03, 0.21]*	-0.04 [-0.13, 0.05]
City_Others	0.08 [-0.04, 0.20]	0.15 [0.03, 0.27]	-0.13 [-0.17,-0.10]***	0.03 [-0.01, 0.07]

Age_1	0.02 [-0.22, 0.25]	0.70 [-0.46, 0.94]***	-1.23 [-1.38, -1.09]***	-0.59 [-0.74, -0.44]***
Age_2	2.60 [2.39, 2.82]***	1.31 [1.08, 1.54]***	3.90 [3.76, 4.03]***	2.55 [2.41, 2.70]***
Ret.Home	2.09 [2.05, 2.14]***	1.36 [1.31, 1.41]***	2.25 [2.22, 2.28]***	1.42 [1.39, 1.46]***
Gender	1.07 [1.04, 1.10]***	0.67 [0.64, 0.70]***	1.32 [1.30, 1.34]***	0.97 [0.95, 0.99]***
Pregnancy	-0.71 [-0.79, -0.64]***	-0.20 [-0.27, -0.12]***	-0.62 [-0.66, -0.57]***	-0.06 [-0.11, -0.02]*
Date_1	-0.05 [-0.18, 0.08]	-0.06 [-0.20, 0.08]	0.06 [0.05, 0.07]***	0.07 [0.06, 0.08]***
Date_2	0.01 [-0.00, 0.02]	0.01 [-0.00, 0.02]	0.00 [-0.00, 0.00]	-0.00 [-0.00, 0.00]
Season_1	-0.02 [-0.04, -0.00]	-0.02 [-0.04, -0.01]*	-0.01 [-0.02, 0.00]	-0.03 [-0.04, -0.02]**
Season_2	0.02 [0.00, 0.04]	0.03 [0.01, 0.05]*	0.02 [0.00, 0.03]*	0.03 [0.01, 0.04]**
Tel-Aviv-Reference	0	0	0	0
City_Haifa	0.48 [0.42, 0.54]***	0.06 [-0.00, 0.13]	0.48 [0.44, 0.52]***	0.07 [0.03, 0.12]*
City_Rishon le Zion	-0.23 [-0.30, -0.16]***	-0.06 [-0.13, 0.02]	-0.29 [-0.34, -0.25]***	-0.10 [-0.15, -0.05]**
City_Holon	-0.01 [-0.08, 0.06]	0.09 [0.02, 0.16]	-0.09 [-0.13, -0.04]**	0.04 [-0.01, 0.08]
City_Jerusalem	0.09 [0.02, 0.16]	0.09 [0.01, 0.16]	0.05 [0.00, 0.10]	0.05 [-0.00, 0.10]
City_Ramat-Hasharon	-0.35 [-0.44, -0.27]***	-0.30 [-0.39, -0.21]***	-0.31 [-0.37, -0.26]***	-0.27 [-0.32, -0.21]***
City_Petah Tiqva	-0.12 [-0.20, -0.04]*	-0.03 [-0.12, 0.05]	-0.14 [-0.19, -0.09]***	-0.07 [-0.12, -0.01]
City_Bat-Yam	0.76 [0.69, 0.82]***	0.39 [0.32, 0.46]***	0.69 [0.65, 0.73]***	0.33 [0.28, 0.37]***
City_Ashdod	0.11 [0.02, 0.19]	0.02 [-0.06, 0.11]	0.05 [-0.00, 0.10]	-0.02 [-0.08, 0.03]
City_Bney Brak	0.02 [-0.06, 0.11]	0.10 [0.01, 0.19]	0.11 [0.06, 0.16]**	0.11 [0.06, 0.17]**
City_Beer-Sheva	-0.09 [-0.17, 0.00]	0.04 [-0.05, 0.13]	-0.07 [-0.12, -0.01]	0.03 [-0.02, 0.09]
City_Netanya	0.25 [0.17, 0.33]***	0.03 [-0.05, 0.11]	0.29 [0.24, 0.34]***	0.05 [-0.00, 0.10]
City_Ramat Gan	-0.03 [-0.12, 0.07]	-0.08 [-0.17, 0.02]	-0.06 [-0.12, -0.01]	-0.06 [-0.12, -0.00]
City_Kiryat Motzkin	0.02 [-0.08, 0.11]	-0.01 [-0.11, 0.09]	-0.01 [-0.08, 0.05]	-0.04 [-0.10, 0.03]
City_Raanana	0.29 [0.19, 0.38]***	0.06 [-0.04, 0.16]	0.24 [0.18, 0.30]***	0.03 [-0.03, 0.10]
City_Kfar Saba	0.07 [-0.04, 0.18]	-0.11 [-0.23, 0.00]	0.11 [0.04, 0.17]*	-0.05 [-0.12, 0.02]
City_Modiin	-0.33 [-0.45, -0.21]***	-0.00 [-0.13, 0.12]	-0.38 [-0.46, -0.29]***	-0.03 [-0.11, 0.06]
City_Ashkelon	0.28 [0.18, 0.38]***	0.24 [0.13, 0.34]**	0.18 [0.11, 0.25]***	0.11 [0.04, 0.18]*
City_Givaataim	-0.13 [-0.26, -0.00]	-0.06 [-0.19, 0.07]	-0.25 [-0.33, -0.17]***	-0.12 [-0.21, -0.04]*
City_Kiryat Ono	0.21 [0.09, 0.32]*	0.09 [-0.03, 0.21]	0.09 [0.01, 0.16]	-0.03 [-0.11, 0.05]
City_Others	-0.04 [-0.09, 0.01]	0.09 [0.04, 0.14]*	-0.13 [-0.16, -0.09]***	0.03 [-0.00, 0.06]

	Cephalexin		Cephalothin	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Age_1	-1.27 [-1.46, -1.08]***	-0.82 [-1.02, -0.63]***	-2.03 [-2.18, -1.88]***	-1.35 [-1.50, -1.19]***
Age_2	3.42 [3.24, 3.61]***	2.43 [2.23, 2.62]***	3.99 [3.84, 4.14]***	2.70 [2.54, 2.85]***
Ret.Home	1.96 [1.91, 2.01]***	1.27 [1.21, 1.32]***	2.05 [2.01, 2.09]***	1.34 [1.29, 1.38]***
Gender	1.21 [1.18, 1.24]***	0.91 [0.88, 0.94]***	1.06 [1.03, 1.08]***	0.78 [0.75, 0.81]***
Pregnancy	-0.40 [-0.46, -0.35]***	0.01 [-0.05, 0.07]	-0.41 [-0.45, -0.37]***	-0.03 [-0.07, 0.02]
Date_1	0.19 [0.01, 0.36]	0.21 [0.03, 0.39]	0.15 [0.13, 0.17]***	0.16 [0.14, 0.18]***
Date_2	-0.01 [-0.02, 0.00]	-0.01 [-0.02, 0.00]	-0.02 [-0.02, -0.02]***	-0.02 [-0.03, -0.02]***
Season_1	-0.01 [-0.02, 0.01]	-0.02 [-0.03, -0.00]	-0.02 [-0.03, -0.00]	-0.02 [-0.03, -0.01]*
Season_2	-0.02 [-0.04, -0.01]*	0.00 [-0.01, 0.02]	0.01 [-0.00, 0.03]	0.02 [0.00, 0.03]
Tel-Aviv-Reference	0	0	0	0
City_Haifa	0.39 [0.34, 0.45]***	0.04 [-0.02, 0.10]	0.41 [0.37, 0.46]***	0.13 [0.08, 0.18]***
City_Rishon le Zion	-0.19 [-0.26, -0.13]***	-0.06 [-0.12, 0.00]	-0.24 [-0.29, -0.19]***	-0.08 [-0.13, -0.03]*
City_Holon	0.03 [-0.03, 0.09]	0.10 [0.03, 0.16]*	-0.14 [-0.19, -0.09]***	-0.02 [-0.07, 0.03]
City_Jerusalem	0.09 [0.02, 0.15]*	0.09 [0.02, 0.15]	-0.02 [-0.07, 0.04]	-0.01 [-0.07, 0.04]
City_Ramat-Hasharon	-0.26 [-0.33, -0.19]***	-0.22 [-0.30, -0.15]***	-0.26 [-0.31, -0.20]***	-0.21 [-0.27, -0.15]***
City_Petah Tiqva	-0.11 [-0.18, -0.04]*	-0.05 [-0.12, 0.03]	-0.12 [-0.17, -0.06]**	-0.07 [-0.13, -0.01]
City_Bat-Yam	0.61 [0.55, 0.67]***	0.32 [0.25, 0.38]***	0.51 [0.46, 0.56]***	0.20 [0.15, 0.26]***
City_Ashdod	0.10 [0.02, 0.17]*	0.00 [-0.07, 0.08]	-0.02 [-0.08, 0.04]	-0.06 [-0.12, 0.00]
City_Bney Brak	0.10 [0.02, 0.17]*	0.17 [0.09, 0.24]**	0.10 [0.04, 0.15]*	0.04 [-0.02, 0.10]
City_Beer-Sheva	-0.09 [-0.17, -0.02]	0.00 [-0.08, 0.08]	-0.02 [-0.08, 0.04]	0.04 [-0.02, 0.10]
City_Netanya	0.25 [0.18, 0.32]***	0.07 [-0.00, 0.14]	0.21 [0.15, 0.27]***	0.04 [-0.02, 0.10]
City_Ramat Gan	-0.01 [-0.09, 0.07]	-0.05 [-0.14, 0.03]	-0.14 [-0.20, -0.07]**	-0.12 [-0.18, -0.06]*
City_Kiryat Motzkin	0.03 [-0.06, 0.11]	0.02 [-0.07, 0.10]	-0.02 [-0.09, 0.05]	-0.02 [-0.10, 0.05]
City_Raanana	0.23 [0.14, 0.31]***	0.08 [-0.01, 0.16]	0.10 [0.03, 0.17]*	-0.05 [-0.13, 0.02]
City_Kfar Saba	0.02 [-0.07, 0.12]	-0.10 [-0.21, -0.00]	0.02 [-0.06, 0.09]	-0.09 [-0.17, -0.01]
City_Modiin	-0.26 [-0.36, -0.15]***	0.01 [-0.10, 0.11]	-0.40 [-0.49, -0.30]***	-0.11 [-0.20, -0.01]
City_Ashkelon	0.18 [0.09, 0.27]**	0.13 [0.04, 0.23]*	-0.02 [-0.10, 0.07]	-0.07 [-0.16, 0.02]
City_Givaataim	-0.13 [-0.25, -0.02]	-0.08 [-0.20, 0.03]	-0.25 [-0.34, -0.16]***	-0.12 [-0.21, -0.03]
City_Kiryat Ono	0.12 [0.01, 0.22]	0.05 [-0.06, 0.15]	-0.05 [-0.13, 0.04]	-0.14 [-0.23, -0.05]*
City_Others	-0.03 [-0.08, 0.01]	0.07 [0.02, 0.11]*	-0.13 [-0.16, -0.09]***	0.01 [-0.02, 0.05]

* P<0.01 ** P<0.0001 *** P<0.000001

Supplementary Table 2: Antimicrobial drugs considered in purchase model.

Group	Name	ATC codes	Counts
1	amoxicillin	J01CA04	1321804
2	amoxicillin/ca	J01CR02	743798
3	cefuroxime	J01DC02	552986
4	cefalexin	J01DB01	507924
5	ciprofloxacin	J01MA02	458447
6	penicillin	J01CE01 J01CE02 J01CE08	363346
7	nitrofurantoin	J01XE01	321099
8	roxithromycin	J01FA06	287591
9	ofloxacin	J01MA01	265931
10	azithromycin	J01FA10	233245
11	trimethoprim/sulfa	J01EE01	232708
12	fosfomycin	J01XX01	175813
13	doxycycline	J01AA02	172187
14	metronidazole	J01XD01	116291
15	minocycline	J01AA08	94004
16	clarithromycin	J01FA09	68021
17	clindamycin	J01FF01	52890
18	erythromycin	J01FA01	48139
19	methenamine	J01XX05	39459
20	ampicillin/inhib	J01CA51 J01CR01	33696
21	cefaclor	J01DC04	22971
22	cloxacillin	J01CF02	21678
23	ceftriaxone	J01DD04	14317
24	cefixime	J01DD08	13944
25	levofloxacin	J01MA12	13944
26	tetracycline	J01AA07	7906
27	ampicillin	J01CA01	4793
28	fusidic acid	J01XC01	4265
29	gentamicin	J01GB03	3087
30	vancomycin	J01XA01	2053
31	amikacin	J01GB06	2000
32	ertapenem	J01DH03	1703
	rifaximin	A07AA11	1645
11	sulfamethoxazol	J01EC01	1567
	cefazolin	J01DB04	1171
	colistin	J01XB01	1156
34	ceftazidime	J01DD02	919
32	meropenem	J01DH02	849
33	piperacillin/inhib	J01CR05	725
	neomycin	A07AA01	704
	spiramycin	J01FA02	643
	clofazimine	J04BA01	630
	nalidixic acid	J01MB02	568
32	imipenem/inhib	J01DH51	412
	norfloxacin	J01MA06	377
	dapsone	J04BA02	224
	moxifloxacin	J01MA14	197
	tobramycin	J01GB01	142
33	piperacillin	J01CA12	127

Author information

Affiliations

Faculty of Biology, Technion–Israel Institute of Technology, Haifa, Israel.

Idan Yelin, Olga Snitser, Ofir Tal & Roy Kishony

Faculty of Computer Science, Technion–Israel Institute of Technology, Haifa, Israel.

Gal Novich & Roy Kishony

Lorry I. Lokey Interdisciplinary Center for Life Sciences & Engineering, Technion–Israel Institute of Technology, Haifa, Israel.

Ofir Tal & Roy Kishony

Maccabitech, Maccabi Healthcare Services, Tel-Aviv, Israel.

Rachel Katz, Gabriel Chodick, Gideon Koren & Varda Shalev

Maccabi Healthcare Services, National Laboratory, Rehovot, Israel.

Miriam Parizade

Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel.

Gabriel Chodick & Varda Shalev

* Correspondence should be addressed to R. Kishony (rkishony@technion.ac.il)

Contributions

V.S. and R.Kishony perceived the study; I.Y., O.S., G.K., V.S. and R.Kishony designed the study; R.Katz, M.P. and V.S. retrieved and interpreted electronic health records; I.Y., O.S., G.N. and R.Kishony analysed the data; I.Y., O. S., G.C., V.S. and R.Kishony interpreted the results; I.Y. and R.Kishony wrote the manuscript with comments from all authors.

Acknowledgements

We thank M. Datta, A. McAdam, G. Priebe and P. Ramesh for thorough reading of the manuscript and important comments. This work was supported in part by US National Institutes of Health grant R01 GM081617 (to RK) and European Research Council FP7 ERC Grant 281891 (to RK).

References

1. Ventola, C. L. The antibiotic resistance crisis: part 1: causes and threats. *P T* **40**, 277–283 (2015).
2. Rossolini, G. M., Arena, F., Pecile, P. & Pollini, S. Update on the antibiotic resistance crisis. *Curr. Opin. Pharmacol.* **18**, 56–60 (2014).
3. Goossens, H., Ferech, M., Vander Stichele, R., Elseviers, M. & ESAC Project Group. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet* **365**, 579–587 (2005).
4. Bronzwaer, S. L. A. M. *et al.* A European study on the relationship between antimicrobial use and antimicrobial resistance. *Emerg. Infect. Dis.* **8**, 278–282 (2002).
5. Costelloe, C., Metcalfe, C., Lovering, A., Mant, D. & Hay, A. D. Effect of antibiotic prescribing in primary care on antimicrobial resistance in individual patients: systematic review and meta-analysis. *BMJ* **340**, c2096 (2010).
6. Kang, C.-I. *et al.* Bloodstream infections caused by antibiotic-resistant gram-negative bacilli: risk factors for mortality and impact of inappropriate initial antimicrobial therapy on outcome. *Antimicrob. Agents Chemother.* **49**, 760–766 (2005).
7. Kumar, A. *et al.* Initiation of inappropriate antimicrobial therapy results in a fivefold reduction of survival in human septic shock. *Chest* **136**, 1237–1248 (2009).
8. Huang, A. M. *et al.* Impact of rapid organism identification via matrix-assisted laser desorption/ionization time-of-flight combined with antimicrobial stewardship team intervention in adult patients with bacteremia and candidemia. *Clin. Infect. Dis.* **57**, 1237–1245 (2013).
9. Stamm, W. E. & Norrby, S. R. Urinary tract infections: disease panorama and challenges. *J.*

- Infect. Dis.* **183 Suppl 1**, S1–4 (2001).
10. Schaeffer, A. J. & Schaeffer, A. J. Infections of the urinary tract. in *Campbell's Urology, Eighth Edition* (Saunders, 2002).
 11. Geerlings, S. E. Clinical Presentations and Epidemiology of Urinary Tract Infections. *Microbiol Spectr* **4**, (2016).
 12. Shapiro, D. J., Hicks, L. A., Pavia, A. T. & Hersh, A. L. Antibiotic prescribing for adults in ambulatory care in the USA, 2007–09. *J. Antimicrob. Chemother.* **69**, 234–240 (2014).
 13. Kahlmeter, G. An international survey of the antimicrobial susceptibility of pathogens from uncomplicated urinary tract infections: the ECOtextperiodcentered SENS Project. *J. Antimicrob. Chemother.* **51**, 69–76 (2003).
 14. Farrell, D. J., Morrissey, I., De Rubeis, D., Robbins, M. & Felmingham, D. A UK multicentre study of the antimicrobial susceptibility of bacterial pathogens causing urinary tract infection. *J. Infect.* **46**, 94–100 (2003).
 15. Foxman, B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Am. J. Med.* **113 Suppl 1A**, 5S–13S (2002).
 16. Flores-Mireles, A. L., Walker, J. N., Caparon, M. & Hultgren, S. J. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* **13**, 269–284 (2015).
 17. Pouwels, K. B. *et al.* Association between use of different antibiotics and trimethoprim resistance: going beyond the obvious crude association. *J. Antimicrob. Chemother.* **73**, 1700–1707 (2018).
 18. Ashkenazi, S., Even-Tov, S., Samra, Z. & Dinari, G. Uropathogens of various childhood populations and their antibiotic susceptibility. *Pediatr. Infect. Dis. J.* **10**, 742–746 (1991).
 19. Kahan, N. R. *et al.* Empiric treatment of uncomplicated urinary tract infection with

fluoroquinolones in older women in Israel: another lost treatment option? *Ann.*

Pharmacother. **40**, 2223–2227 (2006).

20. Hooton, T. M., Besser, R., Foxman, B., Fritsche, T. R. & Nicolle, L. E. Acute uncomplicated cystitis in an era of increasing antibiotic resistance: a proposed approach to empirical therapy. *Clin. Infect. Dis.* **39**, 75–80 (2004).
21. Arslan, H., Azap, O. K., Ergönül, O., Timurkaynak, F. & Urinary Tract Infection Study Group. Risk factors for ciprofloxacin resistance among Escherichia coli strains isolated from community-acquired urinary tract infections in Turkey. *J. Antimicrob. Chemother.* **56**, 914–918 (2005).
22. Ikram, R., Psutka, R., Carter, A. & Priest, P. An outbreak of multi-drug resistant Escherichia coli urinary tract infection in an elderly population: a case-control study of risk factors. *BMC Infect. Dis.* **15**, 224 (2015).
23. Foxman, B. & Brown, P. Epidemiology of urinary tract infections: transmission and risk factors, incidence, and costs. *Infect. Dis. Clin. North Am.* **17**, 227–241 (2003).
24. Tenney, J., Hudson, N., Alnifaidy, H., Li, J. T. C. & Fung, K. H. Risk factors for acquiring multidrug-resistant organisms in urinary tract infections: A systematic literature review. *Saudi Pharmaceutical Journal* (2018). doi:10.1016/j.jsps.2018.02.023
25. Colgan, R., Johnson, J. R., Kuskowski, M. & Gupta, K. Risk factors for trimethoprim-sulfamethoxazole resistance in patients with acute uncomplicated cystitis. *Antimicrob. Agents Chemother.* **52**, 846–851 (2008).
26. Burman, W. J. *et al.* Conventional and molecular epidemiology of trimethoprim-sulfamethoxazole resistance among urinary Escherichia coli isolates. *Am. J. Med.* **115**, 358–364 (2003).
27. Kang, M.-S., Lee, B.-S., Lee, H.-J., Hwang, S.-W. & Han, Z.-A. Prevalence of and Risk

Factors for Multidrug-Resistant Bacteria in Urine Cultures of Spinal Cord Injury Patients.

Ann. Rehabil. Med. **39**, 686–695 (2015).

28. Lee, G., Cho, Y.-H., Shim, B. S. & Lee, S. D. Risk factors for antimicrobial resistance among the *Escherichia coli* strains isolated from Korean patients with acute uncomplicated cystitis: a prospective and nationwide study. *J. Korean Med. Sci.* **25**, 1205–1209 (2010).
29. Johnson, L. *et al.* Emergence of fluoroquinolone resistance in outpatient urinary *Escherichia coli* isolates. *Am. J. Med.* **121**, 876–884 (2008).
30. Paul, M. *et al.* Improving empirical antibiotic treatment using TREAT, a computerized decision support system: cluster randomized trial. *J. Antimicrob. Chemother.* **58**, 1238–1245 (2006).
31. Olesen, S. W., Barnett, M. L., MacFadden, D. R., Lipsitch, M. & Grad, Y. H. Trends in outpatient antibiotic use and prescribing practice among US older adults, 2011-15: observational study. *BMJ* **362**, k3155 (2018).
32. Ena, J., Amador, C., Martinez, C. & Ortiz de la Tabla, V. Risk factors for acquisition of urinary tract infections caused by ciprofloxacin resistant *Escherichia coli*. *J. Urol.* **153**, 117–120 (1995).
33. Brown, P. D., Freeman, A. & Foxman, B. Prevalence and predictors of trimethoprim-sulfamethoxazole resistance among uropathogenic *Escherichia coli* isolates in Michigan. *Clin. Infect. Dis.* **34**, 1061–1066 (2002).
34. Metlay, J. P., Strom, B. L. & Asch, D. A. Prior antimicrobial drug exposure: a risk factor for trimethoprim-sulfamethoxazole-resistant urinary tract infections. *J. Antimicrob. Chemother.* **51**, 963–970 (2003).
35. Wang, A., Daneman, N., Tan, C., Brownstein, J. S. & MacFadden, D. R. Evaluating the Relationship Between Hospital Antibiotic Use and Antibiotic Resistance in Common

- Nosocomial Pathogens. *Infect. Control Hosp. Epidemiol.* **38**, 1457–1463 (2017).
36. Kahlmeter, G. & Mendeny, P. Cross-resistance and associated resistance in 2478 *Escherichia coli* isolates from the Pan-European ECO\textperiodcentered SENS Project surveying the antimicrobial susceptibility of pathogens from uncomplicated urinary tract infections. *J. Antimicrob. Chemother.* **52**, 128–131 (2003).
37. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
38. Lieberman, T. D. *et al.* Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat. Genet.* **43**, 1275–1280 (2011).
39. Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A. & Crook, D. W. Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* **13**, 601–612 (2012).
40. Bradley, P. *et al.* Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat. Commun.* **6**, 10063 (2015).
41. Khoury, M. J. & Ioannidis, J. P. A. Medicine. Big data meets public health. *Science* **346**, 1054–1055 (2014).
42. Beam, A. L. & Kohane, I. S. Big Data and Machine Learning in Health Care. *JAMA* **319**, 1317–1318 (2018).
43. Grad, Y. H. & Lipsitch, M. Epidemiologic data and pathogen genome sequences: a powerful synergy for public health. *Genome Biol.* **15**, 538 (2014).
44. Sandora, T. J., Gerner-Smidt, P. & McAdam, A. J. What's your subtype? The epidemiologic utility of bacterial whole-genome sequencing. *Clin. Chem.* **60**, 586–588 (2014).
45. van der Bij, A. K. *et al.* Clinical breakpoint changes and their impact on surveillance of antimicrobial resistance in *Escherichia coli* causing bacteraemia. *Clin. Microbiol. Infect.* **18**, E466–72 (2012).

46. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **29**, 1189–1232 (2001).