Association of health, nutrition, and socioeconomic variables with global antimicrobial resistance: a modelling study

Association of health, nutrition, and socioeconomic variables with global antimicrobial resistance: a modelling study

Patrick Murigu Kamau Njage, Bram van Bunnik, Patrick Munk, Ana Rita Pinheiro Marques, Frank M Aarestrup

Summary

Background Although antimicrobial use is a key selector for antimicrobial resistance, recent studies have suggested that the ecological context in which antimicrobials are used might provide important factors for the prediction of the emergence and spread of antimicrobial resistance.

Methods We used 1547 variables from the World Bank dataset consisting of socioeconomic, developmental, health, and nutritional indicators; data from a global sewage-based study on antimicrobial resistance (abundance of antimicrobial resistance genes [ARGs]); and data on antimicrobial usage computed from the ECDC database and the IQVIA database. We characterised and built models predicting the global resistome at an antimicrobial class level. We used a generalised linear mixed-effects model to estimate the association between antimicrobial usage and ARG abundance in the sewage samples; a multivariate random forest model to build predictive models for each antimicrobial resistance class and to select the most important variables for ARG abundance; logistic regression models to test the association between the predicted country-level antimicrobial resistance abundance and the country-level proportion of clinical resistant bacterial isolates; finite mixture models to investigate geographical heterogeneities in the abundance of ARGs; and multivariate finite mixture models with covariates to investigate the effect of heterogeneity in the association between the most important variables and the observed ARG abundance across the different country subgroups. We compared our predictions with available clinical phenotypic data from the SENTRY Antimicrobial Surveillance Program from eight antimicrobial classes and 12 genera from 56 countries.

Findings Using antimicrobial use data from between Jan 1, 2016, and Dec 31, 2019, we found that antimicrobial usage was not significantly associated with the global ARG abundance in sewage (p=0·72; incidence rate ratio 1·02 [95% CI 0·92–1·13]), whereas country-specific World Bank's variables explained a large amount of variation. The importance of the World Bank variables differed between antimicrobial classes and countries. Generally, the estimated global ARG abundance was positively associated with the prevalence of clinical phenotypic resistance, with a strong association for bacterial groups in the human gut. The associations between bacterial groups and ARG abundance were positive and significantly different from zero for the aminoglycosides (three of the four of the taxa tested), β-lactam (all the six microbial groups), fluoroquinolones (seven of nine of the microbial groups), glycopeptide (one microbial group tested), folate pathway antagonists (four of five microbial groups), and tetracycline (two of nine microbial groups).

Interpretation Metagenomic analysis of sewage is a robust approach for the surveillance of antimicrobial resistance in pathogens, especially for bacterial groups associated with the human gut. Additional studies on the associations between important socioeconomic, nutritional, and health factors and antimicrobial resistance should consider the variation in these associations between countries and antimicrobial classes.

Funding EU Horizon 2020 and Novo Nordisk Foundation.

Lancet Planet Health 2023; 7: e888–99

Research Group for Genomic Epidemiology, Technical University of Denmark, Lyngby, Denmark (P M K Njage PhD, P Munk PhD, Prof F M Aarestrup PhD); **Roslin Institute, College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, UK** (B van Bunnik PhD); **Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, Special Administrative Region, China** (A R P Marques PhD)

Correspondence to:
Dr Patrick Murigu Kamau Njage, Research Group for Genomic Epidemiology, Technical University of Denmark, Lyngby, 2800, Denmark
panj@food.dtu.dk

Introduction

Antimicrobials are instrumental in treating bacterial infections.[1] However, the development of antimicrobial resistance poses an increasing risk and substantial cost to the economy and health-care systems globally. Estimates of the current economic burden of antimicrobial-resistant bacterial infections are €1·5 billion per year in the EU and US$55 billion in the USA,[2] and there are an estimated 1·27 million deaths directly attributable to antimicrobial resistance.[3] These figures will continue to increase if we do not find ways of reducing the selection and transmission of antimicrobial resistance.

Increases in antimicrobial resistance have mainly been attributed to antimicrobial usage. Although antimicrobial usage might be the main selector for antimicrobial resistance, recent studies have suggested that the ecological context in which antimicrobials are used—namely, different socioeconomic, anthropological, developmental, environmental, health, and nutrition factors—might be associated with and predictive of the increases in antimicrobial resistance.[4–7]

Research in context

Evidence before this study

Antimicrobial resistance is one of the most important threats to human health. Identifying novel control measures is of utmost importance. We searched PubMed for evidence published from database inception to Dec 20, 2019, using the keyword chain (health OR nutrition OR socioeconomic OR anthropological) AND (control OR intervention OR management OR strategies) AND (emergence OR transmission) AND (global OR world) AND (antimicrobial OR antibiotic OR AMR) AND (resistance) NOT ("OneHealth" OR "One Health" OR animal OR livestock). No restriction was placed on language. Increases in antimicrobial resistance have traditionally been attributed to the use of antimicrobial agents. However, although antimicrobial use might be the selector for antimicrobial resistance from the Darwinian perspective of natural selection, recent studies[1–5] have suggested that the ecological context in which antibiotics are used—namely, different socioeconomic, anthropological, developmental, environmental, health, and nutrition factors—might be more important factors for antimicrobial resistance. However, it is not clear how the relative importance of these factors of antimicrobial resistance differs between countries and regions. Additionally, it is not clear whether the predictive role of these factors or the ecological associations might differ between antimicrobial classes.

Added value of this study

We demonstrate that metagenomic analysis of sewage is a robust approach for the surveillance of antimicrobial resistance in pathogens; however, it is especially good for bacterial groups associated with the human gut. We also showed that the abundance of antimicrobial resistance genes (ARGs) in sewage samples is associated with both the pH of the sample and the season in which a sample is taken. Furthermore, we show that the country-level antimicrobial resistance measured using ARG abundance in metagenomics investigations of urban sewage can be reproducibly predicted using socioeconomic, developmental, health, and nutritional indicators. The exact variable and its association were different depending on the antimicrobial class and the country in question, suggesting that the same epidemiological predictors do not apply in all situations or for all countries. These results also show that country-specific tailored interventions should be further investigated in causal studies for addressing the most important antimicrobial resistance at class level for a given country.

Implications of all the available evidence

The global sewage ARG abundance, for most antimicrobial classes, could be used to positively predict the prevalence of resistance in bacteria. The associations were strongly positive for microbial groups associated with human faeces. These data indicate that metagenomic analysis of sewage is not only a robust approach for the surveillance of antimicrobial resistance in all pathogens, but that it can more naturally monitor resistance microbial groups associated with the human gut. Additionally, the influence of pH on ARG abundance is an important consideration when comparing samples between different studies and should be included as a factor in future surveillance efforts. Collecting and publishing good quality metadata is also important. Although additional research is needed, our study suggests that factors for antimicrobial resistance need to be evaluated and tailored to specific countries and antimicrobial classes. Additionally, it could be possible to predict the effect of suggested political interventions on antimicrobial resistance. Our findings also highlight that a thorough understanding of local realities and contexts is central to sustainability-related issues.

The relative importance of these ecological predictors between countries and regions is not known. In a previous study, we found systematic differences in the abundance and diversity of antimicrobial resistance genes (ARGs) in Europe, North America, and Oceania and in Africa, Asia, and South America.[4]

We and others have shown that the global collection of sewage, followed by metagenomic analyses, is a powerful method for obtaining comparable resistome data across many sites and countries.[4,8,9] The use of sewage allows for timely capture of information from a wider community, including hard-to-reach healthy populations, despite the variety of legal and ethical circumstances barring sampling and sharing of human-related samples.[4,9]

For the **World Bank dataset** see https://databank.worldbank.org/home.aspx

In this study, we analyse the relationships at a country and antimicrobial class-specific level between the global sewage resistome (757 samples from 101 countries) and socioeconomic factors. In addition, we computed a global human antimicrobial-usage dataset, spanning from Jan 1, 2016, to Dec 31, 2019, and used it to test the association between antimicrobial usage and ARG abundance. We also compared our predictions to clinical data from central databases covering eight antimicrobial classes and 12 genera from 56 countries.

See **Online** for appendix

Methods

Population-level World Bank and antimicrobial resistance data

We used 1547 variables from the World Bank dataset. The dataset included data on the health, nutrition, and population, and the development indicator data collected between 2015 and 2019 from 264 geographical locations globally, including countries and territories. These data are divided into 20 different multidimensional categories, with each dimension consisting of often highly associated variables (appendix pp 1–2). For each of the

218 dimensions, hierarchical cluster analysis was performed using the Euclidean distance between the predictor variables based on the absolute correlation between variables. A total of 447 variables, consisting of single indicators and indicator representatives from each of the dimensions, were used as machine learning algorithm inputs (appendix pp 1–2).

Antimicrobial resistance data were obtained from a global sewage-based study on antimicrobial resistance,[9] and included metagenomes from 757 samples collected from Jan 1, 2016, to Dec 31, 2019, from 101 countries. Samples of wastewater were collected mainly from treatment plant inlets, as described by Hendriksen and colleagues.[4] Bioinformatic analysis was done as described by Munk and colleagues.[9] Briefly, the antimicrobial resistance data consisted of ARG levels measured in the form of sequencing fragments per kilobase reference gene per million (bacterial 16S rRNA) fragments (FPKM) for 18 antimicrobial classes, which were aminoglycoside, β-lactam, fluoroquinolone, folate pathway antagonist, fosfomycin, glycopeptide, lincosamide, macrolide, oxazolidinone, phenicol, pleuromutilin, polymyxin, quinolone, rifampicin, steroid antibacterial, streptogramin A, streptogramin B, and tetracycline.

To calculate antimicrobial usage variables for each country, we used data from the ECDC database and data from the IQVIA database spanning from Jan 1, 2016, to Dec 31, 2019 (appendix p 2).

Associations between ARGs, antimicrobial usage, and other factors

We estimated the association between antimicrobial usage and ARG abundance in the sewage samples as well as other factors that could influence the distribution of ARGs globally (appendix pp 2–3). We used a generalised linear mixed-effects model, with the counts of ARGs in the different antimicrobial classes as an outcome (Poisson), adjusting for total bacterial 16S rRNA counts and gene length.

Predictive modelling

The World Bank dataset consists of many variables characterised by possible complex interactions, non-linear relations, and mixtures of both discrete and continuous measures. Therefore, we used the machine learning algorithm, a multivariate random forest, to build predictive models and, for each antimicrobial resistance class, to select the most important variables predicting ARG abundance.

The multivariate random forest model was built considering that clustered or correlated antimicrobial resistance classes were responses y_{ik} (i=1,...,n; k=1,...,m) for the 18 antimicrobial resistance classes m and for the same factors for each k. The bootstrap out-of-bag (OOB) error estimates from the random forest models were complemented by a ten-fold cross-validation to obtain the validated predictive accuracy for each antimicrobial resistance class. We did canonical correlation analysis to identify and measure the associations between the predicted and the observed ARG abundance outcomes while considering that the ARG abundance between the antimicrobial classes is intercorrelated (appendix p 4).

We used logistic regression models based on binomial models and logit link function to test the association between the predicted country-level antimicrobial resistance abundance and the country-level proportion of clinical resistant bacterial isolates (appendix pp 2–5). Clinical resistance data were downloaded from the global SENTRY Antimicrobial Surveillance Program covering the current study's time periods between Jan 1, 2016, and Dec 31, 2019. In cases where antimicrobial resistance data for a specific pathogen and antimicrobial class combination were missing for a country, the SENTRY data were complemented with data from WHO's Antimicrobial Resistance Surveillance System, which resulted in data on eight antimicrobial classes and 12 species from 56 countries.

For each of the antimicrobial classes, we selected important factors or variables using the random forest model built for predictive modelling and to indicate which variables might be more strongly associated with ARG abundance. Using OOB sampling, we computed the prediction error (OOB error), which was used to derive the variable importance by assessing the prediction error when each of the variables was used in the model (appendix pp 8–9).[10]

Modelling geographical heterogeneities in the abundance of ARGs

Finite mixture models enabled us to investigate how many of the subpopulations of countries can be differentiated for each antimicrobial class, what underlying antimicrobial resistance categories can be defined for each of the subpopulations, the mean and variance of gene abundance in each subpopulation, and the relative proportion of countries in each subpopulation (appendix pp 9–11). We used the fitted mixture models to compute posterior probabilities and to classify each country into the resistance category or subgroup where it was most likely to belong.[11]

We investigated the effect of heterogeneity in observed ARG abundance across the different country subgroups on the significance, effect size, and direction of the effect of antimicrobial usage and the important socioeconomic, developmental, health, and nutritional factors on ARG abundance. We fitted multivariate finite mixture models with covariates, in which the antimicrobial resistance abundance for the 18 antimicrobial resistance classes were included in each model as responses.

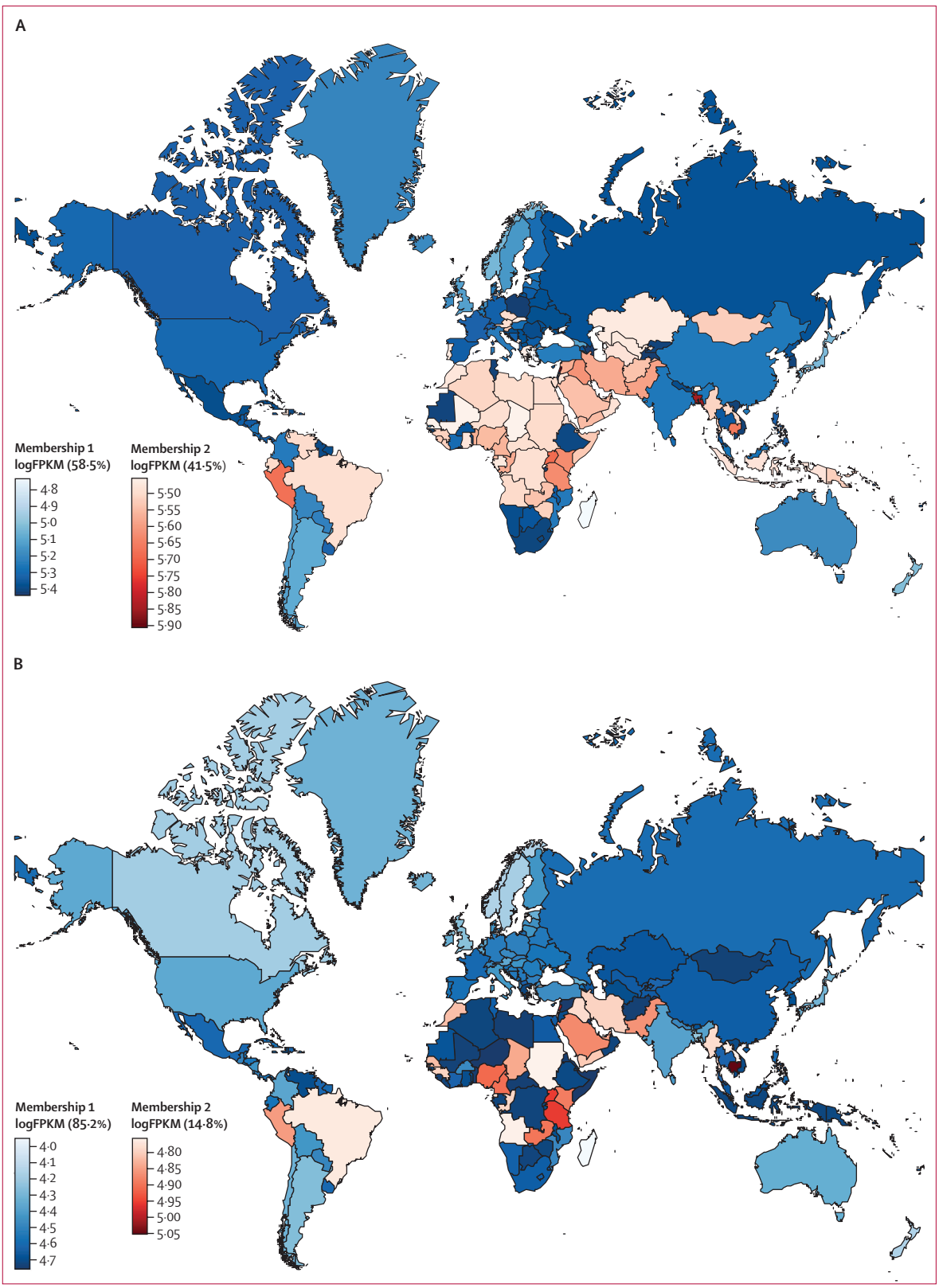
We developed a supporting web-based application to aid the visualisation of data and to reproduce a subset of the results from the statistical analysis using R (version 4.0.1) and the R package, Shiny.[12]

For the **SENTRY Antimicrobial Surveillance Program** see www.jmilabs.com/sentry-surveillance-program

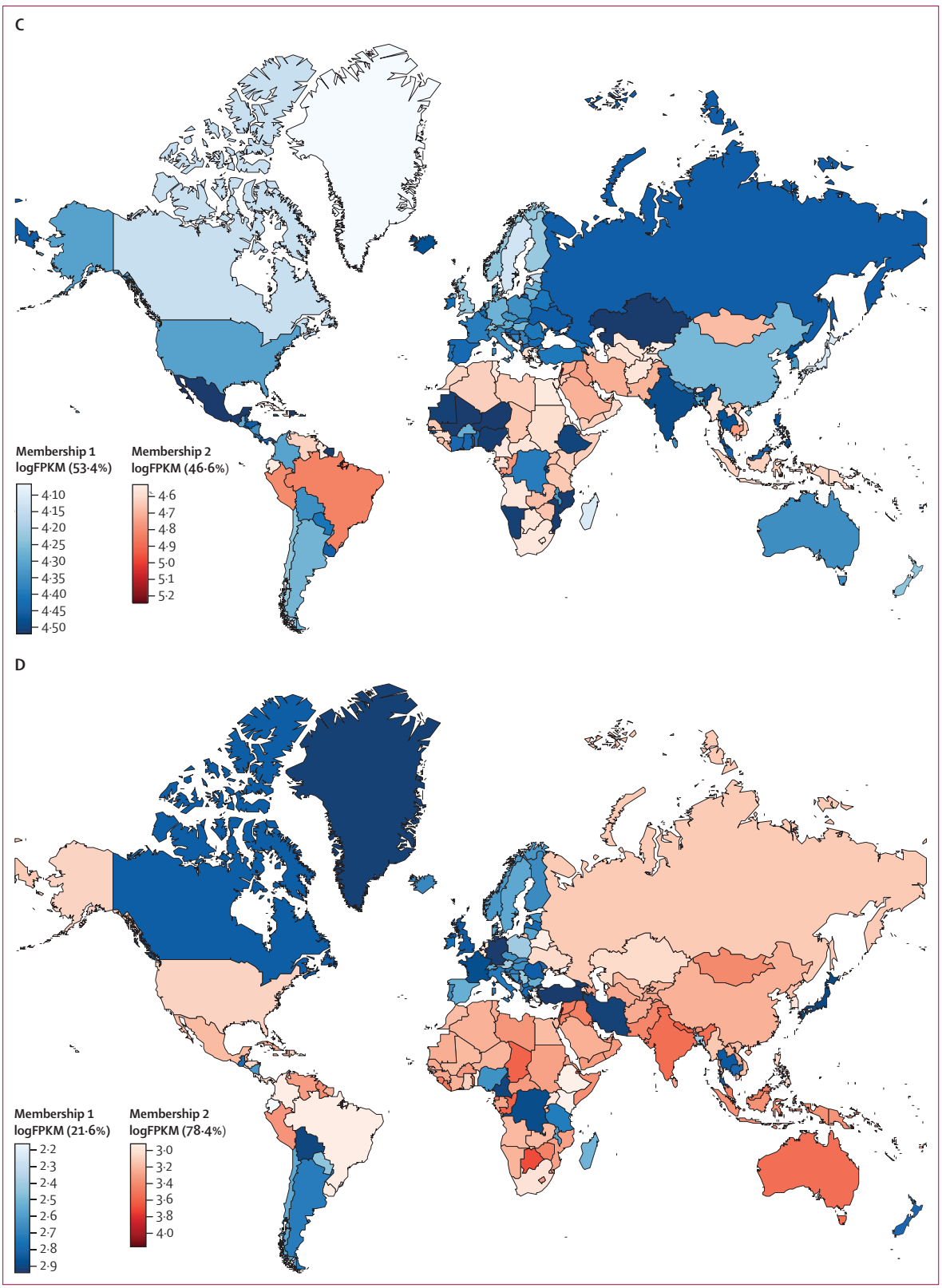
For **WHO's Antimicrobial Resistance Surveillance System** see www.who.int/initiatives/glass

For the **ECDC database** see https://www.ecdc.europa.eu/en/antimicrobial-consumption/surveillance-and-disease-data/database

For the **IQVIA database** see www.iqvia.com



(Figure 1 continues on next page)



(Figure 1 continues on next page)

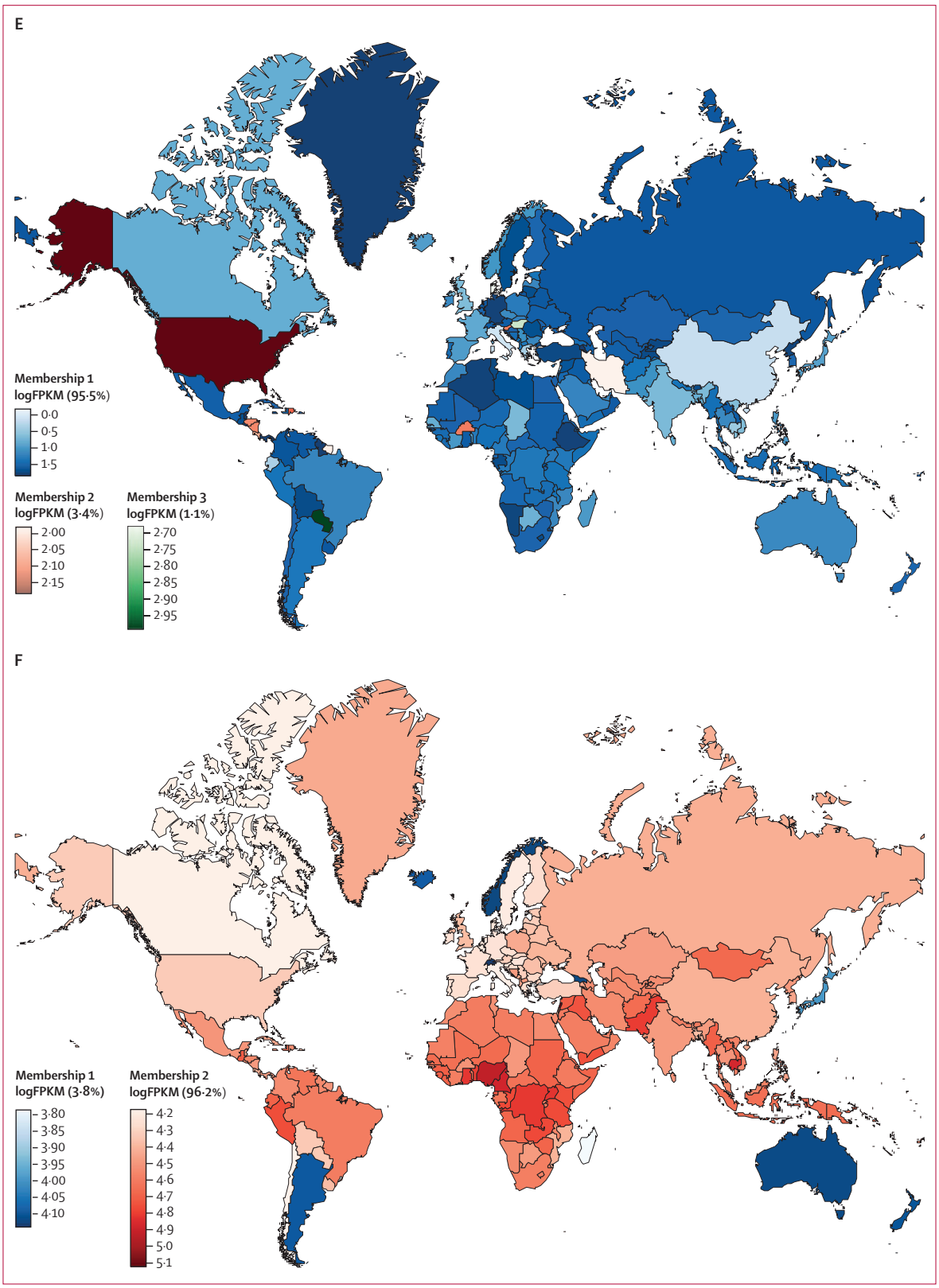


Figure 1: **Predicted global antimicrobial resistance abundance**
(A) Total antimicrobial resistance. (B) Aminoglycoside antimicrobial resistance. (C) β-lactam antimicrobial resistance. (D) Fosfomycin antimicrobial resistance. (E) Glycopeptide antimicrobial resistance. (F) Tetracycline antimicrobial resistance. The maps are coloured according to predicted abundance of antimicrobial resistance by resistance group membership from light (low antimicrobial resistance abundance) to dark (high antimicrobial resistance abundance). Percentages in the legends represent the proportion of countries in each resistance group membership or subpopulation of countries. Data and maps for other classes are presented in the appendix (pp 21–23). FPKM=sequencing fragments per kilobase reference gene per million (bacterial 16S rRNA) fragments.

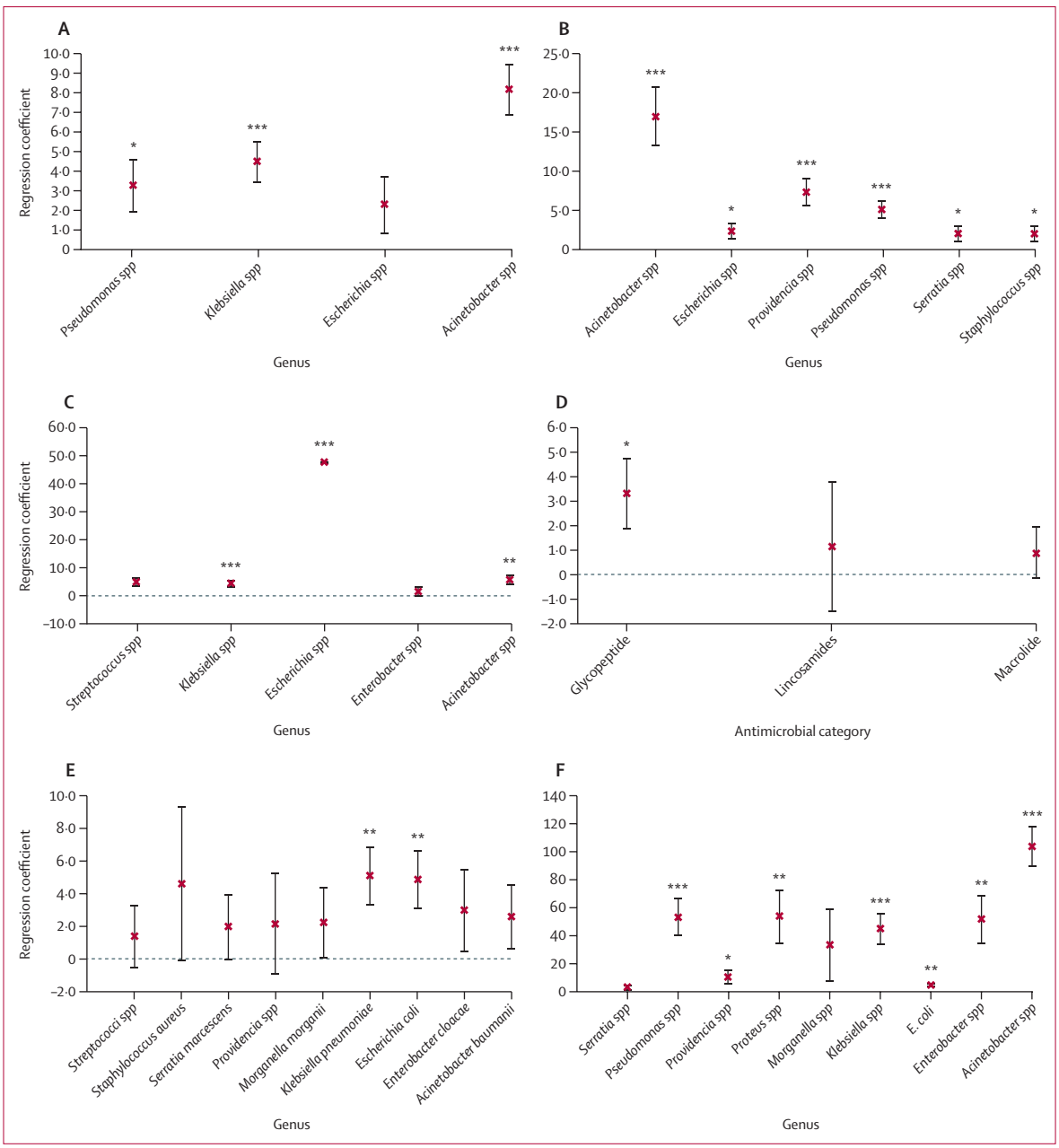


Figure 2: **The association between the predicted antimicrobial resistance abundance and the proportion of clinical resistant bacterial isolates**
(A) Aminoglycoside antimicrobial resistance. (B) β-lactam antimicrobial resistance. (C) Folate pathway antagonists. (D) Glycopeptide, lincosamide, and macrolide antimicrobial resistance to *Streptococci spp.* (E) Tetracycline antimicrobial resistance. (F) Fluoroquinolone antimicrobial resistance. Error bars represent SEs. The regression coefficient represents log odds where the higher the coefficient, the greater abundance of antimicrobial resistance genes. The sum of isolates positive for antimicrobial resistance were for all clinical sites per country and was divided by the total number of isolates tested to obtain a country-level proportion of resistant isolates. *p<0·05. **p<0·01. ***p<0·001.

Role of the funding source
The funders of the study had no role in data collection, data analysis, data interpretation, or writing of the manuscript.

Results
Using data published between Jan 1, 2016, and Dec 31, 2019, we found no significant association between antimicrobial usage (p=0·72; incidence rate ratio [IRR] 1·02 [95% CI 0·92–1·13]) and the abundance of ARGs at the antimicrobial class level, although, for specific antimicrobial classes, we did see a large difference in the fitted intercepts (appendix p 13). Countries that were lower on the human development index scale (HDI) typically had a higher abundance of ARGs in the sample (appendix pp 12, 15). Furthermore, we found an

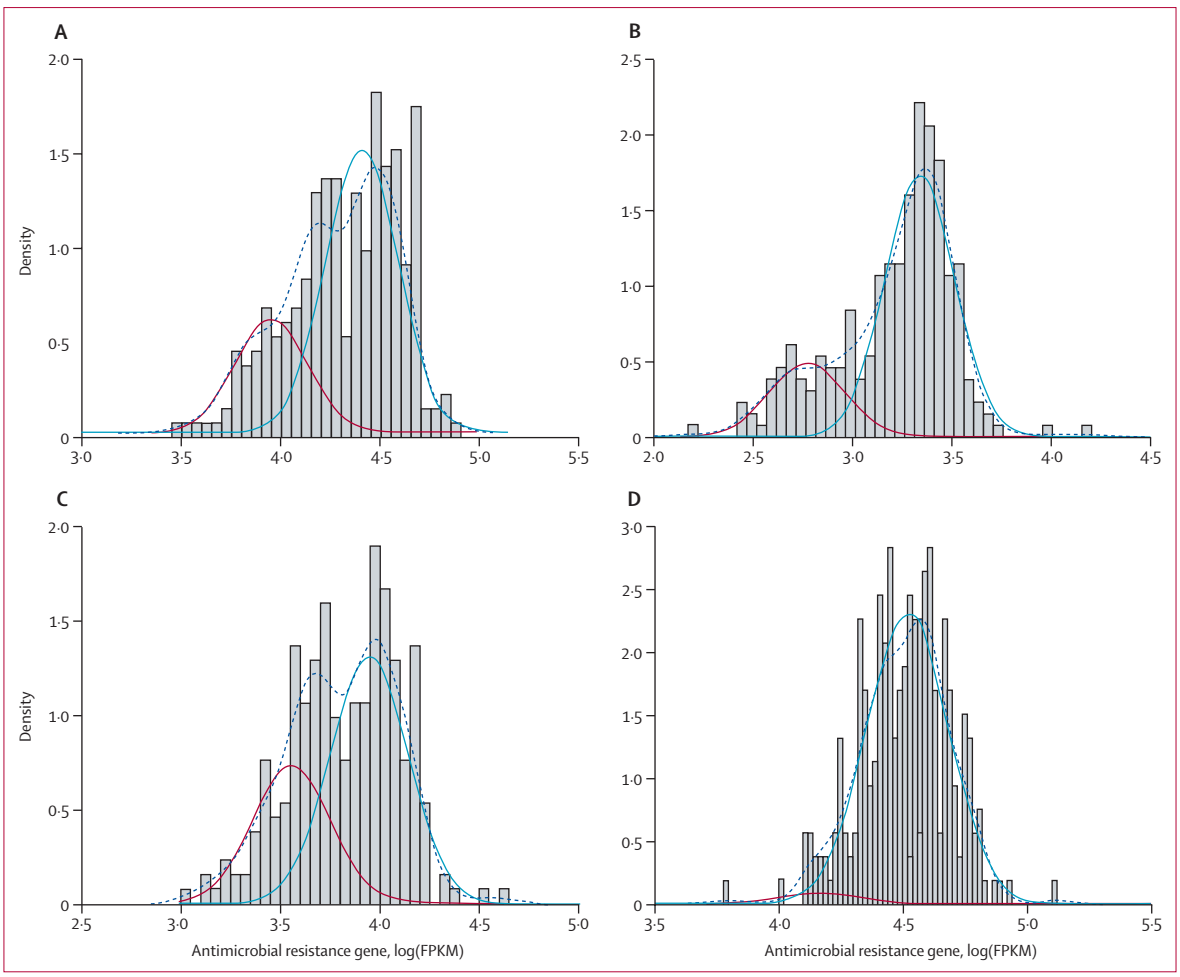


Figure 3: **Distribution of observed antimicrobial resistance abundance**
Illustrated with folate pathway antagonists (A), fosfomycin (B), phenicol (C), and tetracycline (D). The dashed lines indicate a single homogeneous normal distribution. Red and light blue lines indicate the two mixture distributions (subpopulation 1 and 2) underlying the overall (dark blue dashed line) distribution of ARG abundance. Finite mixture models were used to distinguish between subpopulations or subgroups of countries from the overall distribution of the antimicrobial resistance gene abundance data for all countries. FPKM=sequencing fragments per kilobase reference gene per million (bacterial 16S rRNA) fragments.

For the **Shiny web application** see https://patrick-njage.shinyapps.io/GlobalSewageViz/

association of the pH (p=0·0013; IRR 1·24 [95% CI 1·09–1·41]) of the sample with the abundance of ARGs, where lower pH values had markedly lower abundance of ARGs. The results of this analysis also showed that the abundance of resistance genes was markedly higher in dry seasons that fell in temperate spring or summer but not if the dry season fell in a temperate autumn or winter (appendix pp 13–14). All results are summarised in the appendix (pp 12–16).

When we considered total antimicrobial resistance abundance, the five countries with the lowest ARG abundance included Madagascar, New Zealand, Japan, Norway, and Argentina (in order of increasing ARG abundance; figure 1; appendix pp 21–23). Cambodia, Peru, Uganda, Bangladesh, and Lebanon were the five countries with the highest ARG abundance (in order of increasing ARG abundance; figure 1). A detailed and interactive view of the global predictions of the class-level abundance of antimicrobial resistance in 264 countries and territories is presented in the prediction map panel of the Shiny web application.

A detailed summary and discussion of the top and bottom ranking countries for each antimicrobial resistance class are given in the appendix (pp 19–26). Considering class-level ARG abundance, most of the countries that ranked among the five lowest and highest in ARG abundance for one antimicrobial resistance class also ranked among the five lowest or highest for several other antimicrobial resistance classes (appendix p 20). The predictions indicated that the income grouping of the country was not a clear indicator of whether the country belonged to the top or bottom antimicrobial resistance abundance ranks.

The important socioeconomic, health, and nutritional variables factors of class-level antimicrobial resistance abundances are presented in the appendix (pp 28–31).

We identified 78 important socioeconomic, health, and nutritional factors of antimicrobial resistance abundance, considering the predictive importance for all the classes (appendix pp 28–31, 48). Key factors that were common in more than ten antimicrobial resistance classes included: the proportion of the urban population living in slums; the proportion of the population using safely managed drinking water services; the prevalence of anaemia among women who are pregnant; grants awarded to a country excluding those awarded through technical cooperation (current account balance [BoP]; US$); the proportion of rural population practising open defecation; the age dependency ratio (ratio of dependents defined as people younger than 15 years or older than 64 years); time to import (time for customs clearance and inspection procedures); the proportion of people using at least basic drinking water services; mortality caused by road traffic injury (per 100 000 people); the proportion of children younger than 5 years with completed birth registration; the electric power consumption (kWh per person); the proportion of the population at risk of impoverishing expenditure for surgical care; the proportion of children younger than 5 years with completed birth registration in rural settings; the prevalence of anaemia among children younger than 5 years; and the proportion of women in wage employment in the non-agricultural sector (as a proportion of total non-agricultural employment).

The associations between the antimicrobial resistance abundance values from sewage and resistance in clinical pathogens are shown in figure 2 and the appendix (p 32). The associations between the antimicrobial resistance abundance values and clinical resistance were positive in most of the antimicrobial classes for *Escherichia spp*, *Klebsiella spp*, *Acinetobacter spp*, and, in the case of glycopeptides, for *Enterococcus spp*.

Using finite mixture models, we found that, within the main distribution of ARG abundance for the majority of the classes (main population), there were two or more underlying distributions (subpopulations) of ARG abundance (figure 3). Each subpopulation differed by the mean, variance of ARGs, and the relative proportion of countries in each subpopulation (distribution peak height). The subgroups for all countries and territories are presented in the maps of all antimicrobial resistance classes in the appendix (pp 21–23) and can be explored interactively using the Shiny app.

A homogeneous normal distribution would assume an average of 4·32±0·28 log FPKM ARG abundance for folate pathway antagonists (figure 3). However, finite mixture models revealed that the global ARG abundance consisted of the mixture distribution with the two subpopulations. The two subpopulations imply that 68 (26%) of 264 countries and territories were in the lower ARG abundance (3·94±0·17 log FPKM) subgroup, whereas 196 (74%) of 264 countries and territories were in the higher antimicrobial resistance abundance (4·45±0·17 log FPKM) subpopulation of ARG abundance against folate pathway antagonists. Therefore, the population of antimicrobial resistance abundance values can be expressed as a weighted sum of two normal distributions, each with the form $N(\mu,\sigma^2)$, which can be expressed as $Y\,0{\cdot}26N(3{\cdot}94,0{\cdot}17)+0{\cdot}74N(4{\cdot}45,0{\cdot}17)$, where N represents normal distribution with the country proportion 0·26 and 0·27 as weights, μ represents the mean, and σ^2 represents the variance. A similar interpretation can be made from finite mixture models

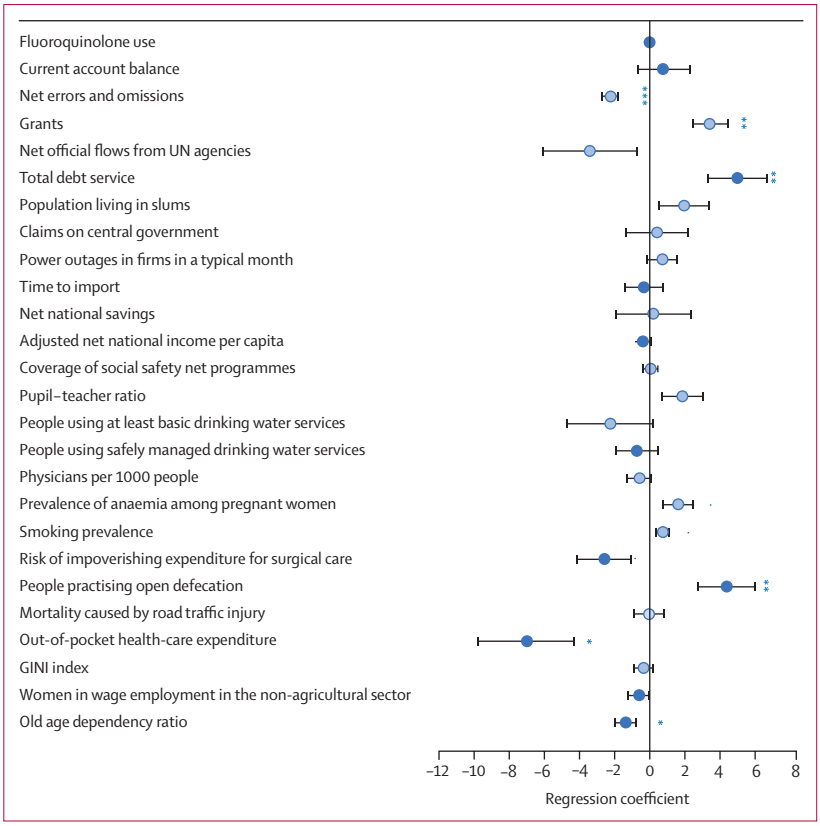


Figure 4: **Generalised linear model coefficients and statistical significance of explanatory variables predicting fluoroquinolone resistance**

Observed antimicrobial resistance abundance was used. Positive coefficients represent an increase in antimicrobial resistance for every unit increase in the variable (risk effect). Negative coefficients represent a decrease in antimicrobial resistance for every unit increase in the variable (protective effect). Showing variables: fluoroquinolone use; current account balance (proportion of gross domestic product); net errors and omissions (current account balance, US$); grants, excluding technical cooperation (current account balance, US$); net official flows from UN agencies (from WHO in US$); total debt service (proportion of gross national income); population living in slums (proportion of urban population); claims on central government (annual growth as a proportion of broad money); power outages in firms in a typical month (number); time to import, border compliance (h); adjusted savings, net national savings proportion of claims on central government); adjusted net national income per person (US$); coverage of social safety net programmes (proportion of population); pupil–teacher ratio, tertiary; people using at least basic drinking water services (proportion of population); people using safely managed drinking water services (proportion of population); physicians (per 1000 people); prevalence of anaemia among women who are pregnant; smoking prevalence (total in people aged 15 years and older); risk of impoverishing expenditure for surgical care (proportion of people at risk); people practising open defecation (proportion of rural population); mortality caused by road traffic injury (per 100 000 people); number of people pushed below the US$1.90 ($2011 purchasing power parity) poverty line by out-of-pocket health care expenditure; GINI index (World Bank estimate); proportion of women in wage employment in the non-agricultural sector (proportion of total non-agricultural employment); and age dependency ratio (ratio of people younger than 15 years or older than 64 years to the working-age population from ages 15–64 years). Significance shows the indicator was significantly different from the alpha value. *p<0·05. **p<0·01. ***p<0·001.

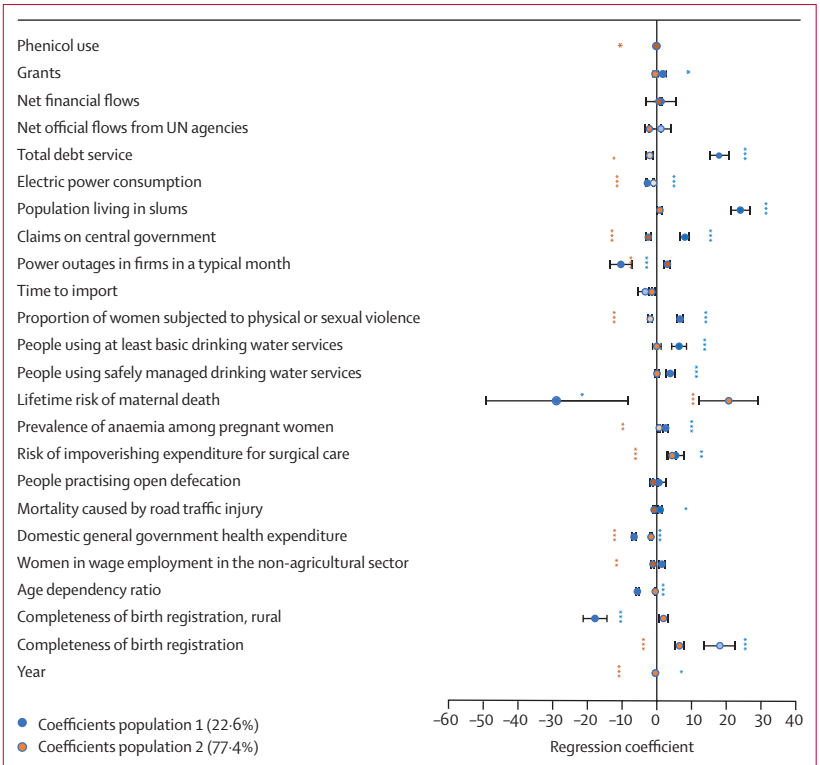


Figure 5: **Multivariate covariate adjusted finite mixture model coefficients and statistical significance of explanatory variables predicting phenicol resistance**

Observed antimicrobial resistance abundance was used. Positive coefficients represent an increase in antimicrobial resistance for every unit increase in the variable (risk effect). Showing variables: phenicol use; grants, excluding technical cooperation (current account balance, US$); net financial flows, repayments of debt balances concessional (net financial flows, US$); net official flows from UN agencies (from WHO in US$); total debt service (proportion of gross national income); electric power consumption (kWh per person); population living in slums (proportion of urban population); claims on central government (annual growth as proportion of broad money); power outages in firms in a typical month (number); time to import, border compliance (h); proportion of women subjected to physical or sexual violence in the last 12 months (women age 15–49 years); people using at least basic drinking water services (proportion of population); people using safely managed drinking water services (proportion of population); lifetime risk of maternal death; prevalence of anaemia among pregnant women; risk of impoverishing expenditure for surgical care (proportion of people at risk); people practising open defecation (proportion of rural population); mortality caused by road traffic injury (per 100 000 people); domestic general government health expenditure (proportion of general government expenditure); women in wage employment in the non-agricultural sector (proportion of total non-agricultural employment); age dependency ratio (ratio of people younger than 15 years or older than 64 years to the working-age population from ages 15–64 years); completeness of birth registration, rural (proportion); completeness of birth registration (proportion); and year. Negative coefficients represent a decrease in antimicrobial resistance for every unit increase in the variable (protective effect). We used covariate-adjusted finite mixture models to derive separate coefficients and statistical tests for each of the two subpopulations of countries that were differentiated for phenicol. Significance shows the indicator was significantly different from the alpha value. *p<0·05. **p<0·01. ***p<0·001.

from the rest of the antimicrobial resistance classes (appendix pp 32–45).

Except for fluoroquinolone resistance (figure 4), the association of explanatory variables differed in magnitude, direction, and statistical significance. Subsequent interpretation of the associations between explanatory variables and antimicrobial resistance should be made individually for each subpopulation of antimicrobial resistance in countries globally. For instance, in the case of phenicol, the two subpopulations of antimicrobial resistance could be represented by two regression models (figure 3, figure 5, appendix pp 34–37). Explanatory variables associated with a large increase in antimicrobial resistance in the first, but not the second subpopulation, included the increase in total debt service; power outages in firms (power outages that establishments have in a typical month); the proportion of lifetime risk of maternal death; and the proportion of people at risk of impoverishing expenditure due to surgical care. However, an increase in the proportion of government expenditure on health from domestic sources as a share of total public expenditure was strongly associated with decreased phenicol antimicrobial resistance for this first subpopulation. When the second subpopulation is considered, phenicol use and the proportion of the urban population living in slums were associated with an increased antimicrobial resistance in the second subpopulation only.

In some instances, explanatory variables in both subpopulations were associated with antimicrobial resistance in the same direction but different magnitude. For instance, marked changes in antimicrobial resistance in both subpopulations were associated with an increase in electric power consumption, the proportion of the children younger than 5 years with completed birth registration, year of sewage sample collection (2017 compared with 2016), and increase in grants awarded to a country excluding those awarded through technical cooperation.

The regression models also enabled us to derive the number of times an explanatory variable strongly predicted higher or reduced antimicrobial resistance for the antimicrobial classes and subpopulations (appendix pp 37–49). The results indicated that 573 (69%) of 831 regression coefficients for explanatory variables from the different antimicrobial resistance classes and country subpopulations were significantly associated either with an increase (294 [51%] of 573 statistically significant coefficients) or decrease (279 [49%] of 573 statistically significant coefficients) in antimicrobial resistance (appendix pp 37–49).

Discussion

In agreement with previous findings,[4,5,13] we found that the available antimicrobial usage data explain only a small proportion of the variation in the global ARG abundance. In recent years, some studies have shown mixed results regarding the association between antimicrobial usage and antimicrobial resistance and some studies show that reducing antimicrobial usage results in reduced antimicrobial resistance,[14,15] whereas other studies suggest that antimicrobial usage only explains some of the variation (at most)[4,5] or find that it has no relation at all.[16] Similar findings have been reported for animal antimicrobial usage.[17] Thus, although we also believe that antimicrobial usage is the main selector for antimicrobial resistance, our results underpin that, in studies investigating interventions to combat antimicrobial resistance, other factors also need to be considered and might even prove to be important.

The World Bank's country-specific health, nutrition, population, and development indicator variables explained a large amount of variation in antimicrobial resistance abundance (16–98%, appendix p 18). Importantly, the amount of variation explained by these factors varied between the different antimicrobial classes.

An area of interest has been the degree to which data from sewage surveillance correlates with antimicrobial resistance in human clinical isolates. Some studies have found a strong correlation when comparing a small number of countries and pathogens.[18–20] In this study, we expanded this research to a comparison of 56 countries and 12 genera, where the phenotypic resistance data were obtained using a central and standardised methodology. The regression coefficients indicated that the global ARG abundance for most of the classes showed a strong positive association with the prevalence of clinical resistance for different microbial groups. A unique trend was the observation from regression coefficients that the associations were positive for microbial groups associated with human faeces. These findings indicate that metagenomic analysis of sewage is not only a robust approach for the surveillance of antimicrobial resistance in all pathogens, but it can specifically be used to monitor the abundance of antimicrobial resistance for microbial groups associated with the human gut. Auguet and colleagues,[21] in a study of three countries, demonstrated that metagenomic analysis of pooled faecal samples has potential in predicting population-level antimicrobial resistance in clinical Enterobacterales isolates.

Another important finding in our study was that the factors and their degree of association with antimicrobial resistance differed between antimicrobial classes and countries. The lower a country was on the HDI scale, the higher the abundance of resistance genes in their sample. This observation could be due to differences in societal contexts between individuals in high-income and low-income countries.[7] In addition, we observed heterogeneities in the data. Additional studies designed to investigate causal effects are needed to confirm the associations inferred from our results and the interventions for further study might need to be tailored to the specific country or antimicrobial class. Additionally, changes in antimicrobial resistance might also be associated with political interventions that influence a country's socioeconomic, developmental, health and nutritional status. We suggest researching country subgroup-specific relationships between the factors and antimicrobial resistance because important factors varied between geographical zones, with protective indicators occurring more in higher income countries and risk factors occurring more in lower income countries (appendix pp 50–54). Our findings suggest that the most frequent and important variables associated with changes in antimicrobial resistance across the classes of antimicrobial resistance were from World Bank variable thematic areas, which closely mirrored UN Sustainable Development Goals (3–8; appendix pp 54–55).[22] These factors might have a crucial role in the post-emergence spread of resistance organisms and genetic elements; however, additional studies are needed to causally ascertain these links.[5,13]

Our results also showed that the abundance of ARGs in sewage samples was associated with both the pH of the sample and the season in which a sample was taken. Although these differences were not sufficiently strong to influence the overall conclusions, the link with seasons is interesting. A possible explanation for this finding could be that, during the wet season, there will be more precipitation and the stream of sewage will have a shorter retention time and less time for sewer-associated bacteria to grow. Additionally, the dependence of the abundance on the pH of the sample is important when comparing samples between different studies and it should be included as a factor in future surveillance efforts. Differences in pH between sources that co-enter sewage might be associated with changes in microbial community composition that might, in turn, affect their ARG profiles, including abundance.[23] Thus, strengthening the importance of collecting (and publishing) good quality metadata.

There were some limitations of our study. Causal inference should not be made from the results of this study, which were based on predictive models and models testing associations. Additionally, confirmatory studies are needed to support causal inference, which are based on disaggregated data with customised study designs to allow collection of antimicrobial usage, antimicrobial resistance, and explanatory data from smaller units. We also did not explicitly study the indirect influence of explanatory variables on antimicrobial resistance and we did not model the co-influence of seasonal changes in sewage temperature and changes in hydraulic conditions on antimicrobial resistance.

In conclusion, metagenomic analysis of sewage is a robust approach for the surveillance of antimicrobial resistance in all pathogens and can more naturally monitor antimicrobial resistance for bacterial groups associated with the human gut. The abundance of antimicrobial resistance can be reproducibly predicted using socioeconomic data from the World Bank. Importantly, the degree of correlation of the explanatory variables associated with antimicrobial resistance differs between countries and antimicrobial classes. Future studies on interventions to improve antimicrobial resistance situations should consider that the association between the factors and antimicrobial resistance differs between countries and antimicrobial classes. This study suggests that the associations of mitigation measures and microbial resistance can vary widely depending on the country-specific direct and indirect interplay between these factors.

Contributors

PMKN and FMA contributed to the conceptualisation of the study, the investigation, and the writing of the original draft. FMA contributed to funding acquisition, data acquisition, and project administration. PMKN, BvB, PM, and ARPM contributed to methodology, data acquisition, curation, formal analysis, visualisation, interpretation, and the review and editing of the manuscript. All authors had full access to all the data in the study, approved the final version of the submitted manuscript, and agreed to be accountable for all aspects of the work.

Declaration of interests

All authors declare no competing interests.

Data sharing

The raw sequencing data (FASTQ) used in this study were deposited in the European Nucleotide Archive and are available without restrictions under the project numbers PRJEB40798, PRJEB40816, PRJEB40815, PRJEB27621, and ERP015409. Exact sample data, and experiment and run accessions are available from Munk and colleagues. Antimicrobial usage and World Bank datasets used for our analyses are freely available from the sources cited in the Methods section of this article. Detailed outline of the models, mathematical notation and specification, and citations with vignettes containing the open access software packages used are provided in the appendix.

Acknowledgments

This study was supported by the EU's Horizon H2020 grant VEO (874735) and Novo Nordisk Foundation (NNF16OC0021856). We thank Lynda Wainaina (Department of Mathematics, University of Padova, Padova, Italy) for her help in conducting canonical correlation analysis.

References

1 WHO. Antimicrobial resistance: global report on surveillance. Geneva: World Health Organization, 2014.
2 Mobarki N, Almerabi B, Hattan A. Antibiotic resistance crisis. *IJMDC* 2019; **3:** 561–64.
3 Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* 2022; **399:** 629–55.
4 Hendriksen RS, Munk P, Njage P, et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat Commun* 2019; **10:** 1124.
5 Collignon P, Beggs JJ, Walsh TR, Gandra S, Laxminarayan R. Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. *Lancet Planet Health* 2018; **2:** e398–405.
6 Allel K, García P, Labarca J, Munita JM, Rendic M, Undurraga EA. Socioeconomic factors associated with antimicrobial resistance of *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Escherichia coli* in Chilean hospitals (2008–2017). *Rev Panam Salud Publica* 2020; **44:** e30.
7 Miller-Petrie M, Gelband H. Socioeconomics, antimicrobial use and antimicrobial resistance. AMR Control. Aug 2, 2017. http://resistancecontrol.info/2017/socioeconomics-antimicrobial-use-and-antimicrobial-resistance/ (accessed June 14, 2022).
8 Su JQ, An XL, Li B, et al. Metagenomics of urban sewage identifies an extensively shared antibiotic resistome in China. *Microbiome* 2017; **5:** 84.
9 Munk P, Brinch C, Møller FD, et al. Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance. *Nat Commun* 2022; **13:** 7251.
10 Ishwaran H, Lu M. Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Stat Med* 2019; **38:** 558–82.
11 Schlattmann P. Medical applications of finite mixture models, statistics for biology and health. Berlin: Springer, 2009.
12 Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. Shiny: web application framework for R. R package version 0.11.1. 2015. https://cran.r-project.org/web/packages/shiny/index.html (accessed Sept 22, 2023).
13 Collignon P, Beggs JJ. Socioeconomic enablers for contagion: factors impelling the antimicrobial resistance epidemic. *Antibiotics (Basel)* 2019; **8:** 86.
14 Aarestrup FM, Seyfarth AM, Emborg HD, Pedersen K, Hendriksen RS, Bager F. Effect of abolishment of the use of antimicrobial agents for growth promotion on occurrence of antimicrobial resistance in fecal enterococci from food animals in Denmark. *Antimicrob Agents Chemother* 2001; **45:** 2054–59.
15 Dutil L, Irwin R, Finley R, et al. Ceftiofur resistance in Salmonella enterica serovar Heidelberg from chicken meat and humans, Canada. *Emerg Infect Dis* 2010; **16:** 48–54.
16 Bell BG, Schellevis F, Stobberingh E, Goossens H, Pringle M. A systematic review and meta-analysis of the effects of antibiotic consumption on antibiotic resistance. *BMC Infect Dis* 2014; **14:** 13.
17 Birkegård AC, Halasa T, Græsbøll K, Clasen J, Folkesson A, Toft N. Association between selected antimicrobial resistance genes and antimicrobial exposure in Danish pig farms. *Sci Rep* 2017; **7:** 1–8.
18 Pärnänen KMM, Narciso-da-Rocha C, Kneis D, et al. Antibiotic resistance in European wastewater treatment plants mirrors the pattern of clinical antibiotic resistance prevalence. *Sci Adv* 2019; **5:** eaau9124.
19 Hutinel M, Huijbers PMC, Fick J, Åhrén C, Larsson DGJ, Flach CF. Population-level surveillance of antibiotic resistance in *Escherichia coli* through sewage analysis. *Euro Surveill* 2019; **24:** 1800497.
20 Karkman A, Berglund F, Flach CF, Kristiansson E, Larsson DGJ. Predicting clinical resistance prevalence using sewage metagenomic data. *Commun Biol* 2020; **3:** 711.
21 Auguet OT, Niehus R, Gweon HS, et al. Population-level faecal metagenomic profiling as a tool to predict antimicrobial resistance in *Enterobacterales* isolates causing invasive infections: an exploratory study across Cambodia, Kenya, and the UK. *EClinicalMedicine* 2021; **36:** 100910.
22 UN. Sustainable Development Goals. Sept 20, 2023. www.un.org/sustainabledevelopment/sustainable-development-goals/ (accessed Jan 24, 2023).
23 Ju F, Beck K, Yin X, et al. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. *ISME J* 2019; **13:** 346–60.