UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics

Masterthesis

Persistence of antimicrobial resistance in E. coli after exposure to penicillins or fluoroquinolones

Yilikal Tesfaye Haile

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :
Mevrouw Robin BRUYNDONCKX

SUPERVISOR :
Dr. Boudewijn CATRY

2017
2018

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics

Masterthesis

Persistence of antimicrobial resistance in E. coli after exposure to penicillins or fluoroquinolones

Yilikal Tesfaye Haile

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :

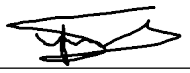
Mevrouw Robin BRUYNDONCKX

SUPERVISOR :

Dr. Boudewijn CATRY

Declaration of Authorship

I, Yilikal Tesfaye Haile, declare that this thesis titled, 'Persistence of antimicrobial resistance in *E. coli* after exposure to penicillins or Fluoroquinolone' and the work presented in it is my own. I confirm that this work submitted for assessment is my own and is expressed in my own words. Any uses made within it of the works of other authors in any form (e.g., ideas, equations, figures, text, tables, programs) are properly acknowledged at any point of their use. A list of the references employed is included.

Signed: 

Date: June 2018

Abstract

Background: Antimicrobial resistance is a worldwide public health issue. It implies the ability of microorganisms (like bacteria, viruses and some parasites) to stop an antimicrobial from working against it. Accordingly, standard treatments become inefficient and infections persist in the body, increasing the risk of spread to others.

Objectives: To investigate to what extent and for how long commensal bacteria, such as *E .coli* retrieved from the digestive tract (gut), are affected by treatment with broad-spectrum aminopenicillins and fluoroquinolones. And, if yes, whether this impact is different from the effect of other types of antibiotics tested in the laboratory.

Methods and Materials: The dataset was collected in a multicenter cohort study during the periods 2010-2012 and patient prescription obtained from national reimbursement data(collected by the intermutualistic agency - IMA Brussels). Antimicrobial resistance index (ARI) was used as a response variable, which is computed by aggregating resistance to multiple antimicrobial tested for each *E. coli* isolate. Multiple logistic regression and Generalized Estimating Equations were used to analyze the data.

Results: The total study population (sample size) was 236 samples from 120 patients. For the first outcome of interest, ARI_1, the estimated odds of resistance when the treatment administered aminopenicillins equal 0.62 times the estimated odds of resistance when the treatment provided fluoroquinolones. Moreover, there was an interaction effect between gender of the patient and log(time) on the log odds of resistance. Concerning the second outcome of interest, ARI_2, the odds ratio of treatment is 0.96 for the male patient while 0.56 for the female patient. For a unit change log(time), the log odds of resistance decrease by -0.26 for male patients and increase by 0.039 for female patients.

Conclusions: The *E. coli* isolates from the digestive tract (gut) had developed resistance to broad-spectrum antibiotics namely aminopenicillins or fluoroquinolones, and the impact is not the same from the effect of other types of antibiotics tested in the laboratory. The impact of number days between the last antimicrobial prescription and sampling date on the probability of resistance depends on the gender of the patient.

Key Words: Antimicrobial resistance; *Escherichia coli (E. coli)*; Aminopenicillins; Fluoroquinolones; Generalized Estimating Equations; Multiple logistic regression.

Acknowledgements

First and foremost, I would like to offer my special thanks to God for blessing me much more than I deserve. I am always looking to him in my life.

Secondly, I would like to give thanks to my advisor's Dr.Robin Bruyndonckx and Dr. Boudewijn Catry for their patient guidance, enthusiastic encouragement, and useful critiques of this research work. Their willingness to give their time so generously has been very much appreciated. I would also like gratitude VLIR-UOS scholarship for the financial support that gives me the chance to join the University of Hasselt in 2016/17 and all my professors at Hasselt University.

Finally, I must express my very profound gratitude to my dad Mr.Tesfaye Haile and to my mom Mrs.Etete Kebede, who have provided me through moral and emotional support in my life. Many Thanks! I am grateful to my brothers Bekalu Tesfaye and Nibret Tesfaye, to my sister Lamirot Tesfaye, and to all my parents and friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of working this thesis. This accomplishment would not have been possible without them.

Thank you!
Yilikal Tesfaye Haile

Contents

List of Figures

List of Tables

Abbreviations

ARI	**A**ntimicrobial **R**esistance **I**ndex
ATC	**A**natomical **T**herapeutic **C**hemical
DDD	**D**efined **D**aily **D**ose
EARS-Net	**E**uropean **A**ntimicrobial **R**esistance **S**urveillance **N**etwork
ECDC	**E**uropean **C**entre for **D**isease Prevention and **C**ontrol
E.Coli	***E**scherichia coli*
EEA	**E**uropean **E**conomic **A**rea
EU	**E**uropean **U**nion
EDA	**E**xploratory **D**ata **A**nalysis
GEE	**G**eneralized **E**stimating **E**quations
IMA	**I**ntermutualistic **A**gency
ICU	**I**ntensive **C**are **U**nit
J01	**A**ntibacterials for systemic use
J01C	**B**eta-Lactem antibacterials, penicillins
J01M	**Q**uinolone Antibacterials
US	**U**nited **S**tates
WHO	**W**orld **H**ealth **O**rganization

Chapter 1

1. Introduction

1.1 Background of the study

1.1.1 Antibiotics

Antibiotics have become one of the most important medical interventions needed for the development of complex medical approaches such as cutting edge surgical procedures, solid organ transplantation, and management of patients with cancer, among others [Lin et al., 2015]. They are molecules that kill or stop the growth of, microorganisms, including both bacteria and fungi. The origin of antibiotics is natural(fungal), semi-synthetic (chemically -altered natural compound), or synthetic (chemically designed in the lab). Most classes of antibiotics, including the betalactam antibiotics, tetracyclines, aminoglycosides, and macrolides are originally derived from natural sources and were then further chemically modified to confer better properties on the formulation and pharmacology of the drug. However, some important classes of antibiotics (including the sulphonamides, the quinolones, and the oxazolidinones) are man-made, originating totally from synthetic chemical operations [Wright et al., 2014].

Antibiotics are usually further classified based on their molecular structure and/or mode of action. Antibiotics operate by inhibiting crucial life sustaining processes in the organism: the cell wall synthesis (e.g. betalactams), the protein and nucleic acid synthesis, the synthesis of DNA (e.g. fluoroquinolones), RNA and ribosome (e.g. macrolides,

aminoglycosides). Finally, classification of antibiotics based on the mode of action depends on inhibiting the life sustaining process of the organisms [Etebu and Arikekpar, 2016]. They can demonstrate both bactericidal (kill bacteria) and bacteriostatic (stop the growth of bacteria) effects that rely on both the concentration and the organism of interest.

1.1.2 Antimicrobial resistance

Antimicrobial resistance is a global public health problem. It implies that microorganisms acquire the ability to survive exposure to antibiotics. Accordingly, standard treatments become inefficient and infections persist in the body, increasing the risk of therapy failure and spread to others. Bacteria use two major genetic strategies to adapt to the antibiotic "attack" resulting in survival despite the presence of the antimicrobial molecule: i) mutations in a gene often associated with a subset of bacterial cells derived from a susceptible population that *vertically* spread to daughter colonies, and ii) acquisition of foreign DNA coding for resistance determinants through *horizontal* gene transfer [Munita and Arias, 2016]. After every initiation of antimicrobial therapy, antimicrobial resistance can occur, which can persist long after the treatment has been finished (persistence of resistance).

1.1.3 *Escherichia coli*(*E.coli*) Bacteria

Escherichia coli are a large and diverse group of bacteria. They are normally found in the lower intestines of warm blooded organisms. Most strains of *E. coli* are harmless, others can be a cause for disorders of the normal health status. The latter type are the most frequent cause of community and hospital-acquired urinary tract infections (including infections of the kidney), a cause of bloodstream infection at all ages, a cause of meningitis in neonates, and one of the leading causative agents of food borne infections worldwide. Moreover, it is associated with intra-abdominal infections such as peritonitis and with skin and soft tissue infections due to multiple microorganisms [WHO et al., 2014].

1.2 Objective of the study

The main objective of this study was to investigate to what extent and for how long commensal bacteria, such as *E .coli* retrieved from the gut, are affected by treatment with broad spectrum penicillins (aminopenicillines) and fluoroquinolones. And, if yes, whether this impact is different from the effect of other types of antibiotics tested in the laboratory.

1.3 Literature review

Resistance in commensal *E. coli* from healthy community members has been first demonstrated over 40 years ago and many more recent studies have highlighted the high or increasing incidence of antibiotic resistance in commensal *E. coli* from healthy children and adults across many countries [Bailey et al., 2010]. Several factors appear to increase the number of antibiotic-resistant of commensal bacteria in the digestive tract(gut). Over prescription of antibiotics for symptoms that in many cases may not be caused by bacteria, improper treatment regimens or incompliance to the prescription, a lack of public knowledge about antibiotics, which also may lead to overuse among both patients and animals are the main factors [Fair and Tor, 2014].

Antibiotics are highly prescribed in the ICU, where patients are continuously exposed to antibiotic therapy because of their increased risk of infection. An investigation in the US shows that there was a strong association between total fluoroquinolones use in the communities surrounding the hospitals and fluoroquinolone-resistant *E. coli* in hospitalized patients [MacDougall et al., 2005]. Antimicrobial resistance surveillance in Belgium indicates that for invasive *E. coli* from the blood and cerebrospinal fluid the percentage of resistance to fluoroquinolones was 27.3 and 24.5 in 2015 and 2016 respectively [Thomas and Karl, 2017].

Resistance to fluoroquinolones, one of the most widely used medicines for the treatment of urinary tract infections, is intensely widespread. According to WHO surveillance global report on antimicrobial resistance, *E. coli* resistance to fluoroquinolones exceeds 50% in five of the world health organization regions [WHO et al., 2014].

Aminopenicillins are recommended for the wide range of infections that made ampicillin one of the most commonly prescribed agents, in particular for urinary and respiratory infections, as well as for gastrointestinal infections [Finch et al., 2010]. However, the occurrence of penicillin resistance becomes a challenge in clinical practice. The report of European Antimicrobial Resistance Surveillance Network (EARS-Net) showed that *E.coli* resistance to aminopenicillins in the EU/EEA was 57.2% in 2015 [ECDC, 2017].

Antimicrobial resistance is one of the greatest threats to human health. Studies show that the yearly impact of resistant infections is estimated to be a 20 billion dollar and an over 1.6 billion euro increase on the health care costs and 8 million and 2.5 million additional hospital days in the United States (US) and European Union (EU), respectively [Fair and Tor, 2014]. Increased antimicrobial resistance is the cause of severe infections, increase length of hospital stays, complications and increased mortality [Llor and Bjerrum, 2014].

There are a number of strategies that can be used to lower antimicrobial resistance. Antibiotic concentrations have an effect on the emergence of resistant in pathogens. Optimization of antibiotic regimens on the basis of pharmacokinetic and pharmacodynamic principles could play a role in the reduction of antibiotic resistance [Martin et al., 2010]. In general, to minimize antibiotics resistance the Infectious Diseases Society of America and the Society for Health care Epidemiology developed a guideline for antimicrobial stewardship teams in hospitals [Dellit et al., 2007]. In addition, effectively diagnosing infections using high-quality cultures, decreases the duration of therapy when possible, use of appropriate aggressive antimicrobial dosing based on the type and severity of infection and body weight when applicable are recommended to minimize antimicrobial resistance [Martin et al., 2010].

Chapter 2

Methods and Materials

2.1 Data description

The dataset of this study was collected in a multicenter cohort study during the periods 2010-2012 and patient prescriptions were obtained from national reimbursement data (collected by the intermutualistic agency – IMA, Brussels). In all volunteered laboratories *E. coli* isolated from faeces underwent susceptibility testing for various antimicrobial agents. In this thesis, among the antimicrobial tested in the laboratories, the focus is only on the antibiotic with ATC code 'J01 – antibacterials for systemic use'.

2.1.1 Laboratory dataset

The dataset obtained from the laboratories contain the variables patient id, sample date, lab id, microorganism isolated such as *E. coli*, antimicrobials tested, susceptibility test results, and sociodemographic characteristics of the patients such as gender, death/alive, and year of birth. Antibacterial susceptibility testing was conducted in five different laboratories. The results from antibacterial susceptibility tests generally classified bacteria as resistant, intermediate resistant, or susceptible to the individual antibacterial drug tested. Susceptible means bacteria can't grow at predefined in vitro drug concentrations. This means the antibiotic is effective against the bacteria. Resistance means the bacteria were able to grow under these in vitro conditions. The intermediate category

is the buffer zone between susceptible and resistant test result, to compensate for laboratory and pharmacodynamics variance. Based on an intermediate resistant result, the antibiotics still can be effective working against the target pathogens provided a high or higher than normal dosage of the drug will be achieved at the site of infection. So, prescribing such antibiotics as treatment may not be appropriate.

For the purpose of the present study, the antibacterial susceptibility test result is categorized into two groups; resistance and intermediate results are together considered as a resistance group. This means the outcome variable result has two categories, namely resistance (result = 1) and susceptible (result = 0).

2.1.2 Prescription dataset

On the prescription dataset, the following variables were measured patient id, delivery date, ATC code of prescribed antibiotics, defined daily dose (n_DDD), administration route (oral or parenteral), and place of prescription (ambulatory or hospital). The defined daily dose (DDD) defined by world health organization (WHO) according to the Anatomical Therapeutic Chemical (ATC) system was used for the classification of active ingredients. Herein drugs are stratified according to the organ or system on which they act as well as based upon therapeutic, pharmacological, and chemical properties [Hutchinson et al., 2004]. Different antimicrobials could have been prescribed for one single patient for a wide variety of pathogens. Among antimicrobials prescribed, the systemic antibacterial classes of aminopenicillins (J01C) and fluoroquinolones (J01M) antibiotics were extracted from the prescription dataset for this study. These antibiotics were assumed to be used to treat infections caused by bacteria in the upper respiratory tract. So, this study includes only patients assumed to have been infected by a bacterium and does not focus on the target pathogens at the site of infection, but on the reservoir of resistance in the digestive tract.

2.1.3 Final dataset

Prior to analysis, data cleaning and data manipulation was carried out. The information obtained from the prescription dataset and the laboratory test result dataset were combined. Additional variables were computed like time, number of prescription before

sample, and proportion of orally administered agents before sample for further analysis. An antibiotic prescribed multiple times within seven consecutive days and by the same route of administration was considered as one prescription. and their total defined daily doses (DDDs) were summed up. The variable time was defined as the number of days between sampling date and prescription date. Prescriptions that were prescribed for less than two days before the sampling date were excluded from the dataset. This was in order to ascertain a default buffer period to all patients to ensure that they started taking the prescribed antibiotics before the sampling. For further analysis, the closest prescription date to each sampling date was used as a value of time variable per patient. The variable number prescription before sample was also computed and defined as the count of the total number of prescribed antibiotics per patient before sample. Finally, the total study population (sample size) was 236 samples from 120 patients. The list of predictor variables are presented in Table 2.1.

2.1.4 Variables in the study

Response variable

Antimicrobial resistance index (ARI) was used as a response variable, which is computed by aggregating resistance to multiple antimicrobial tested for each individual sample. As mentioned above the objectives of this study was to assess the impact of (amino)penicillins and/or fluoroquinolones treatment working against the bacteria such as E. coli in the digestive tract(gut) and whether this impact is different from the effect of other types of antibiotics tested in the laboratories. To achieve these objectives ARI was computed from the laboratory dataset in two different contexts. The first set was the proportion of antibacterial susceptibility tests from antibiotics fluoroquinolones (J01M) and (amino)penicillins (J01C) tested in the laboratory. The second exhaustive setting is that the proportion of resistance test result from all antibiotics 'J01' tested in the laboratory. The respective definitions are as follows:

$$ARI\_1 = \frac{Number\ of\ Resistance\ test\ (result = 1)}{Total\ number\ of\ antibiotics\ tested\ (only\ 'J01M'\ and\ 'J01C')}$$

and

$$ARI\_2 = \frac{Number\ of\ Resistance\ test\ (result\ = 1)}{Total\ number\ of\ antibiotics\ tested\ ('J01')}$$

Therefore, the two response variables are considered separately in this study.

Explanatory variables

Overview of explanatory variables included in the final dataset were presented in the Table 2.1.

TABLE 2.1: List of Predictor variables in the study

Variable	Type	Units of Measurement
Place of prescription	Categorical	Ambulatory or hospital
Age	Continuous	Years
DDD	Continuous	Abolute values (mg/ 70 kg)
Numbers of Prescription per sample	Discrete	count of prescription before sample
Treatment	Categorical	Fluoroquinolones or Aminopenicillins
Proportion of oral administration before sample	Continuous	-
Laboratory id	Categorical	105, 108, 109, 110, or 111
Gender	Categorical	Female or Male
Patient living status	Categorical	died or alive
Time	Discrete	number of days

NB: DDD defined daily dose.

2.2 Exploratory data analysis

Examining descriptively the structure of the variables in the dataset is a crucial step, before making inferences from the data. An approach to visualizing the structure of the data is called exploratory data analysis(EDA). EDA involves looking at the data in many ways to gain insights and incorporating domain knowledge about the data into the analyses. Basic numerical summary and graphs are common methods of exploring the data.

2.3 Multiple Binomial Logistic regression

Most often multiple logistic regression is used to model a binary (0,1) outcome variable based on more than one other variable, called predictors. The binary variable being modeled is generally referred to as the response variable or the dependent variable. The predictors can be a continuous or categorical type. Logistic regression does not require many of the principal assumptions of linear regression models that are based on ordinary least squares estimation method. Moreover, logistic regression can handle non-linear relationships between the response and explanatory variables because it applies a non-linear log transformation of the linear regression. In other words, it models the logit-transformed probability as a linear relationship with the predictor variables [Hosmer Jr et al., 2013]. More officially, let y be the binary response variable signifying success or failure with 1 or 0 and π be the probability of y to be 1. Let x_1, x_2, ..., x_p be a set predictor variables. Next, the multiple logistic regression models with logit link function is presented as follows, which models the log odds of probability of "success" as a function of explanatory variables :-

$$logit(\pi) = log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p$$

where π is the probability of success, X's are the explanatory variables, and β's are the coefficients of these explanatory variables. This model considers that observations are independent and the response follows a binomial distribution. Maximum likelihood estimation method was used to estimate the coefficient of explanatory variables. The likelihood equation for a logistic regression model does not have a closed form solution,

so the maximum likelihood estimates are obtained by using iterative algorithms such as Newton-Raphson (NR), or Iteratively re-weighted least squares (IRWLS) [Czepiel, 2002].

2.4 Generalized Estimating Equations

The generalized estimating equations (GEEs) methodology, first introduced by Liang and Zeger [Liang and Zeger, 1986], enables us to analyze correlated data. These data sets can arise from clustering, in which measurements are taken on subjects who share a common characteristic. This commonality of subjects within cluster violates the basic assumption of likelihood theory. Recall that likelihood method require that each observation, or record, in the model is independent of the others. The resulting correlation within the group biases the standard errors, and even the parameter estimates, of the model [Hilbe, 2009].

The GEEs approach is a widely used estimation method for correlated data. GEEs technique is a direct extension of quasi-likelihood theory from cross-sectional data to clustered measurements. The within-subject correlations among the repeated measures are taken into account by using a working correlation structure which describes the pattern of association amongst the observations that are within each cluster. Liang and Zeger claimed that the GEEs is robust to misspecifications of the working correlation structure [Liang and Zeger, 1986]. Thus, whether or not the working correlation structure is correct, point estimates and standard errors are asymptotically correct. Though in GEE, the working correlation is thought of as being a nuisance parameter and estimates from the model are meant to be robust even if the incorrect correlation structure is specified, it is good scientific practice to carefully select the correlation structure of our data [McGahan, 2017].

Moreover, to account for implicit clustering or correlation, there are several statistical modeling methods that can be used for a binary outcome. Typically these can be grouped into two classes; marginal or population averaged approach, and conditional or cluster specific approach. A GEEs approach is often used for the population averaged approach. The GEE and Cluster specific models resolve the problem of correlated observations but incorporate it into the model differently. With respect to interpretation, the fundamental

distinction between the two methods is the GEEs method models the average response effect across all clusters i.e. it models the marginal expectations of the outcome. In a cluster-specific model, the response effect is specific for a given cluster [Hilbe, 2009]. Recalling to the goal of this study, we want to estimate the average effect over the entire patient group rather than estimate the effect for a patient. That is why we choose GEEs technique in this study.

The GEEs method doesn't specify the complete multivariate distribution, instead, it defines the marginal distributions and the correlation structure [Agresti, 1996]. Due to this reason, there is no likelihood function. In this sense, the GEEs method is a multivariate type of quasi-likelihood method. So, its estimates are not ML estimates. In the case of clustered data, the GEEs method computationally simpler than maximum likelihood estimation method. But, it has constraints. It is known that it does not have a likelihood function, so likelihood ratio test is not applicable for checking fit, comparing model, and conducting inference about parameters. In place of that Wald statistics have been used, which is based on the approximate normality of the estimators together with their estimated covariance matrix. The price to pay for such inference to be decent is sufficient sample size. If not, the empirically based standard errors tend to underestimate the true ones.

2.5 Model selection

Model selection is an essential part of any statistical analysis. The purpose of model selection is to choose a sparse model that adequately explains the data. Several mechanisms for selecting the best model have been advised in the literature. Backward selection, forward selection, and stepwise selection are among the most popular and widespread techniques. They all provide systematic ways of exploring through models, where at each step new models are obtained by adding or deleting one variable from the models at the previous stages.

Forward variable selection method begins with nothing but an intercept in the model. At each step test the addition of each variable using a chosen criterion and add the variable (if any) that improves the model the most. This procedure repeatedly executed until none improve the model. Backward selection method starts with all possible variables in

the model and tests the deletion of each variable using a chosen criterion. The method at each step excludes the variable (if any) that improves the model the most by being deleted and repeat until no further improvement is possible. Stepwise Selection technique is a combination of the previous two methods and test at each stop for variables to be added or removed [Neter et al., 1996].

In both forward selection and backward selection, once a variable has been acted upon, that decision cannot be modified. Hence, a variable that was eliminated at some point during a backward procedure, for example, will never be permitted back into the model. In contrast, in the case of stepwise selection, at each step variables are contemplated concerning incorporation and concerning exclusion. In other words, a variable might be included in an early stage, but taken out later; or, a variable that was taken out of the model might be allowed back in [Kadane and Lazar, 2004]. For this reason, we will focus on the stepwise approach to variable selection.

2.5.1 Stepwise model building

The stepwise approach is an enormously popular technique for model building. It is the procedure for selection and deletion of variables from the model based on the statistical algorithm that checks for the significance of the coefficients of the variable. In statistical term, a significant variable is defined as a variable that contributes to a big change in the log-likelihood of the model. The outcomes of interest of this study were Antimicrobial resistance indexes (ARI_1 and ARI_2) which is the proportion of resistance per sample per patient. In more detail, the samples were clustered within the patient. The predictor variables included in the final dataset were proportion of oral route of administration before sample (Prop_OralAdmin), the number of days between last prescription date and sampling date pre-sample (log(time)), number of prescription before sample (Presc), defined daily dose before sample (DDD), age of patient, gender of patient, place of prescription, the patient living status up to the end of the study year, and the laboratory samples were tested (lab_id). Among the five laboratories, approximately 95% of the samples were tested in one laboratory (Table A.3). This implies that the number of samples tested in the other laboratory were insufficient to compare the variability in the outcome of interest due to variation in the laboratory. Therefore, the variable lab_id was not included in the variable selection. Medical literature recommends that it is possible

to include clinically and intuitively important variables into the model regardless of their statistical significance. The reason behind this approach is to control the confounding effect. Based on this fact log(time) variable was kept fixed in the selection procedure.

To identify important predictors which have a relationship with the outcome of interest stepwise selection procedure was implemented. Once the main effects were identified possible pairwise interactions were also checked among the variables in the model. The choice of level of significance (α) to decide the importance of variables is a key factor in using stepwise logistic regression. Hosmer and Lemeshow recommended using a value from 0.15 to 0.2 for the significance level for entry into the model and the significance level for staying in the model are the best choice [Hosmer Jr et al., 2013]. In this study, 0.1 for entry in the model and 0.15 for a stay in the model were used. Since the samples were repeatedly taken from the same patient, this indicates the samples from the same patient might be correlated. To account such correlations GEEs technique was implemented on the final model selected. Pan proposed a criterion based on quasi-likelihood, named QIC, which can be used to find an acceptable working correlation structure for a given model [Pan, 2001]. Autoregressive, compound symmetry and independent working correlation assumption were checked. Finally, multiple logistic regression models were built for the two outcomes separately. In the models built generalized estimating equations with autoregressive working correlation structure was adopted.

2.6 Software

The model fitting process was conducted using SAS 9.4 software and 5% level of significance was used in all cases to test the significance of the parameter estimate.

Chapter 3

Result

3.1 Exploratory data analysis

3.1.1 Descriptive statistics of qualitative and quantitative variables

TABLE 3.1: The total number of *E. coli* samples taken per patient in the Bacterial Susceptibility test, 2010-2012.

Samples per patient	Number of Patient	Percent
1	72	60 %
2	20	16.67 %
3	12	10 %
4	7	5.83 %
5	2	1.67 %
6	2	1.67 %
7	2	1.67 %
8	3	2.5 %

In Table 3.1 the number of samples taken per patient was presented. The maximum and the minimum number of samples taken per patient is eight and one respectively. This shows that number of observation per patient varies from one to eight. Among the total number of patients in the study, 60% of the patients were sampled only once during the study period.

TABLE 3.2: The total number of penicillins and fluoroquinolones antibiotics administrated with sociodemographic characteristics of the patients included in this study, 2010-2012.

		Treatment			
		Fluoroquinolones	Penicillins	Total	Percent
Gender	Male	40	99	139	58.9 %
	Female	35	62	97	41.1 %
Place of prescription	Ambulatory	53	121	174	73.7 %
	Hospital	22	40	62	26.3 %
Patient living status	died	22	26	48	20.34 %
	alive	53	135	188	79.66 %

Table 3.2 shows that the number of patients received the two treatments together with the gender of the patients, the place of prescription was given and whether the patient alive or not staged. From the result, we observe that 161(68.22%) patients received aminopenicillins and 75(31.78%) patients received fluoroquinolones. This reveals that the group of antibiotics that are commonly used to treat gastrointestinal infection caused by *E. coli* prescribed more frequently. Furthermore, 139(59%) patients were male and 97(41.1%) patients were female. Finally, 62(26.3%) patients were administered the treatments in the hospital and 174 (73.7%)) patients were an ambulatory patient. This shows that a high number of patients not confined to bed in the hospital. Most of the patients have not died during the study year, 188(79.66%).

TABLE 3.3: The mean and standard deviation of patient characteristics measured in 2010-2012.

Variable	Mean	Std Dev	Lower quartile	Upper quartile
DDD	58.61	96.52	17.5	166.3
Presc	4.37	4.84	2	16
Age	35.09	26.27	6	56
Log(Time)	3.35	1.47	2.3	4.46
Prop_OralAdmin	0.85	0.29	0.89	1

NB: DDD: defined daily dose, Prop_OralAdmin: Proportion Of Oral route Administration before sample, and Presc: Number of prescription before sample.

Table 3.3 display the mean, the standard deviation, the lower quartile and the upper quartile of quantitative predictor variables. The mean age of the patient was 35.09 year with standard deviation 26.27. The standard deviation of the age of the patient was high this shows the presence of considerable variation in the age of patients. The average of the log of number of days between closest prescription date to sampling date was 3.35 with standard deviation 1.47. Moreover, the mean of a defined daily dose was 58.61 with standard deviation 96.52, and the average of the number of prescription before sample was 4.37 with standard deviation 4.84. Lastly, on average the proportion oral route of administration was 0.85. This shows that on average most of the antibiotics were taken through oral route administration.

TABLE 3.4: A table shows summary statistic of patient characteristics for male and female patients separately, 2010-2012.

	Female				Male			
Variable	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
DDD	45.49	69.35	.87	374.74	67.77	110.96	0.75	778.5
Presc	4.41	4.64	1	25	4.35	4.96	1	36
Age	30.47	26.14	0	77	38.31	25.97	0	83
log(Time)	3.51	1.53	1.09	6.52	3.24	1.43	1.09	5.91
Prop_OralAdmin	0.89	0.24	0	1	0.82	0.32	0	1
ARI_1	0.45	0.25	0	0.81	0.49	0.25	0	0.89
ARI_2	0.53	0.27	0	1	0.62	0.28	0	1

NB: DDD: defined daily dose, Prop_OralAdmin: Proportion Of Oral route Administration before sample, Presc: number of prescription before sample, ARI_1: Antimicrobial resistance index from 'J01C' and 'J01M' antibiotics tested, and ARI_2: Antimicrobial resistance index form 'J01 antibacterial for systemic use' tested in the laboratory.

Summary statistics of quantitative variables separately presented in Table 3.4 for male and female patients. The mean age of male patients was higher than female patients (38.31, 30.47). On average the number prescription before sample and proportion of oral route administration before sample for male and female patients approximately equal. The mean and the standard deviation of defined daily dose for male patients higher than female patients (67.77 and 45.49 the mean of male and female, 110.96 and 69.35 the standard deviation male and female respectively). Furthermore, there was not so much difference in the mean of antimicrobial resistance index one (ARI_1) between male

and female patients (0.49,0.46). Likewise, the mean of antimicrobial resistance index two(ARI_2) of the male and female patients has not differed (0.62,0.53).

TABLE 3.5: A table shows summary statistic of patient characteristics for hospital and ambulatory patients separately, 2010-2012.

	Hospital				Ambulatory			
Variable	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
DDD	63.30	91.71	0.75	457.97	56.94	98.38	2	778.5
Presc	5.19	6.41	1	36	4.08	4.09	1	31
Age	39.15	29.04	0	81	33.64	25.14	0	83
Log(Time)	2.08	1.12	1.09	4.97	3.80	1.32	1.09	6.52
Prop_oralAdmin	0.49	0.35	0	1	0.98	0.06	0.5	1
ARI_1	0.49	0.22	0	0.88	0.47	0.26	0	0.89
ARI_2	0.61	0.23	0	0.83	0.57	0.29	0	1

NB: DDD: defined daily dose, Prop_OralAdmin: Proportion Of Oral route Administration before sample, Presc: number of prescription before sample, ARI_1: Antimicrobial resistance index from 'J01C' and 'J01M' antibiotics tested, and ARI_2: Antimicrobial resistance index form 'J01 antibacterial for systemic use' tested in the laboratory.

On average the defined daily dose(DDD) and the number of prescription before sample higher for hospital patients. The proportion of oral route of administration on average remarkably higher for an ambulatory patient than hospital patient as shown in Table 3.5. This shows most of the patient consumed the prescribed drug through oral route of administration (Table in the Appendix A.1) , and around 73.7% of the patients were an ambulatory patient (Table 3.2) indicates a possible correlation. Finally, on average the Antimicrobial resistance indexes (ARI_1 and ARI_2) were nearly similar for hospital and an ambulatory patient.

TABLE 3.6: A table shows summary statistics of patient characteristics for the patients died and alive up to the end of the study year, 2010-2012.

	Died				Alive			
Variable	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
DDD	105.43	173.79	0.75	778.5	46.66	58.39	0.87	457.98
Presc	6.29	6.37	1	31	3.88	4.22	1	36
Age	48.94	23.91	0	75	31.55	25.73	0	83
log(time)	2.96	1.36	1.09	6.20	3.45	1.49	1.09	6.52
Prop_oralAdmin	0.73	0.32	0	1	0.88	0.28	0	1
ARI_1	0.51	0.23	0.1	0.88	0.47	0.25	0	0.89
ARI_2	0.63	0.22	0	0.83	0.57	0.29	0	1

NB: DDD: defined daily dose, Prop_OralAdmin: Proportion Of Oral route Administration before sample, Presc: number of prescription before sample, ARI_1: Antimicrobial resistance index from 'J01C' and 'J01M' antibiotics tested, and ARI_2: Antimicrobial resistance index form 'J01 antibacterial for systemic use' tested in the laboratory.

Roughly 80% of the patients were not died up to the end of the study year (Table 3.2). On average the patients died were older than those patients alive up to the end of the study year (48.94, 31.55). The mean and the standard deviation of a defined daily dose of patient died up to the end of the study year were 105.43 and 173.79 and those alive to the end of the study year were 46.66 and 58.39 respectively. There was nearly resemblance in Antimicrobial resistance indexes between patients not died and died up to the end of the study year.

Table 3.7 shows that the patients who administrated with fluoroquinolones the mean of a defined daily dose was 54.95 with a standard deviation of 71.09 and the mean number of prescription before sample 3.83 with standard deviation 4.13. On average the age the patients (mean = 48 and sd = 18.38) who treated with fluoroquinolones was higher than those patient treated with aminopenicillins (mean = 29 and sd = 27.33). The mean of the number of days between closest prescription date and sampling date was 2.93 for the patients prescribed with fluoroquinolones and 3.54 for patients treated with aminopenicillins.

TABLE 3.7: A table shows summary statistics of patient characteristics for the patients treated with Fluoroquinolones and aminopencillins separately, 2010-2012.

	Fluoroquinolones				Aminopenicillins			
Variable	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
DDD	54.95	71.09	2.5	500.59	60.32	106.47	0.75	778.5
Presc	3.83	4.13	1	31	4.63	5.1	1	36
AGE	47.79	18.38	7	81	29.17	27.33	0	83
log(time)	2.93	1.4	1.09	6.52	3.54	1.47	1.09	6.38
Prop_OralAdmin	0.92	0.18	0.33	1	0.82	0.33	0	1
ARI_1	0.53	0.25	0	0.88	0.45	0.24	0	0.89
ARI_2	0.66	0.27	0	1	0.55	0.28	0	1

NB: DDD: defined daily dose, Prop_OralAdmin: Proportion Of Oral route Administration before sample, Presc: number of prescription before sample, ARI_1: Antimicrobial resistance index from 'J01C' and 'J01M' antibiotics tested, and ARI_2: Antimicrobial resistance index form 'J01 antibacterial for systemic use' tested in the laboratory.

TABLE 3.8: Pearson Correlation Coefficients between continuous predictor variable

	DDD	Presc	Age	log(time)
DDD	1	0.802	0.243	-0.122
		<.0001	0.0002	0.0618
Presc		1	0.062	-0.184
			0.342	0.0047
AGE			1	-0.224
				0.0005
log(time)				1

NB: DDD: defined daily dose. Presc: number of prescription before sample.

Statistically, there was a significant high positive association between the defined daily dose and the number of prescription per sample $(0.802, 0.0001)$, defined daily dose and age of the patient $(0.243, 0.0002)$. Likewise, a significant negative association between the number of prescription per sample and log of time $(-0.184, 0.0047)$, and the age of the patient and log of time (-0.224,0.0005). In contrary, there was no significant correlation between defined daily dose and a log of time $(-0.122, < 0.0618)$, and the number of prescription per sample and the age of the patient $(0.062, 0.342)$. Even if some pairwise

correlations between predictor variables were significant, their magnitude showed that the association was not strong except the correlation between defined daily dose and the number of prescription per sample. Therefore, the continuous predictor variables included in the model building process were defined daily dose, age, and log(time). To visualize the correlation graphically a scatter plot between continuous predictor variables plotted (Appendix B.1).

The distribution of Antimicrobial resistance index two (ARI_2) did not differ dramatically between male and female patients (Appendix B.2). On the other hand, the Figure in the appendix B.2 on the left suggests the presence of a slight difference in the distribution of Antimicrobial resistance index one (ARI_1) between male and female patients. The mean value of antimicrobial resistance index one(ARI_1) has not differed too much between male and female patients.

Box plot in the appendix B.3 reveals the distribution of antimicrobial resistance indexes. It seems that there was no substantial difference in the distribution of antimicrobial resistance index one (ARI_1) between patients in the hospital and ambulatory. Similarly, the variability of the antimicrobial resistance index two(ARI_2) of an ambulatory patient compared with the hospital patient has not differed.

Figure in the appendix B.4 depicts that the mean, the median as well as the distribution of antimicrobial resistance indexes did not differ between the patients died and alive up to the end of the study year.

3.2 Multiple Binomial logistic regression

3.2.1 Ordinary logistic regression analysis

The first outcome of interest of this study was ARI_1, the proportion of antibacterial susceptibility tests from antibiotics fluoroquinolones (J01M) and aminopenicillins (J01C) tested in the laboratory. In our analysis, after model building procedure we started fitting the ordinary logistic regression, which modeled the probability of resistance given predictors. A logistic regression needs each observation to be independent. In other words, the observations did not evolve from a design of repeated measurements. Thus,

we ignored the correlation between samples within a patient and assumed that they were independent. The result of the ordinary logistic regression model presented in Table 3.9.

TABLE 3.9: Maximum Likelihood Parameter Estimates, standard error, Odds ratio, and p-values of the coefficients of predictor variables when the outcome is Antimicrobial resistance index one (ARI_1)

Parameter	Estimate	Std error	OR	p-value
Intercept	0.528	0.270	1.69	0.0509
age	0.004	0.002	1.00	0.0593
treatment	-0.745	0.189	0.47	<.0001
gender	0.439	0.311	1.55	0.1719
log(time)	-0.024	0.056	0.98	0.6680
treatment*gender	0.653	0.251	1.92	0.0092
log(time)*gender	-0.162	0.078	0.85	0.0374

Reference group: Treatment - Fluoroquinolone; Gender - Female.

The coefficients of predictors estimated through the method of Maximum Likelihood Estimation. From the result, the age of the patient has a statistically insignificant effect on the log odds of resistance (p-value = 0.0593). However, there was a significant interaction effect between treatment and gender of the patient on the log odds of resistance. It means that the impact of treatment on the log odds of resistance depends on the gender of the patients. Similarly, the variables log(time) and gender have a statistically significant interaction effect on the log odds of resistance.

The second outcome of interest of this study was ARI_2; the exhaustive setting is that the proportion of resistance test result from all antibiotics 'J01' tested in the laboratory. The result of the fitted ordinary logistic regression model displayed in Table 3.10. The result suggests that statistically significant interaction effect between the variables treatment and gender on the log odds of resistance with p-value equals to 0.0022. Likewise, the variable log(time) has also an interaction effect on the log odds of resistance with the gender of the patient.

TABLE 3.10: Maximum Likelihood Parameter Estimates, standard error, Odds ratio, and p-values of the coefficients of predictor variables when the outcome is Antimicrobial resistance index two (ARI_2)

Parameter	Estimate	Std error	OR	p -value
Intercept	0.109	0.133	1.11	0.4101
log(time)	-0.027	0.032	0.97	0.3958
treatment	-0.506	0.100	0.60	<.0001
gender	0.378	0.169	1.46	0.0262
log(time)*gender	-0.130	0.044	0.88	0.0029
treatment*gender	0.419	0.137	1.52	0.0022

Reference group: treatment - fluoroquinolone; gender - female;

The main effects, as well as interaction effects, are significant. These might be due to underestimation of the standard error of the parameter estimates. The fact that not accounting for the correlation between samples within the patient might have led to the underestimation of the standard error of the parameter estimate. Finally, GEEs technique which considers such correlation into account used and the details described in the following section.

3.2.2 Generalized Estimating Equations

It is highly likely that samples from the same patient will be more like one another than samples from different patients. This means that the assumptions of independent and identically distribution of the observed outcome would violate. Therefore, ordinary logistic regression might not fit the data well in this situation. Moreover, if there is a failure to take the clustering into account, then the standard errors of parameter estimate would be underestimated and hence inflate the type one error. The favored method in epidemiological studies which is most often used to analyze correlated data, GEEs technique, was adopted in this study [Liang and Zeger, 1986]. Previously fitted logistics regression models were used to implement GEEs technique. In addition, backward model selection procedure was performed in order to have the most important variables in the final model.

To account the correlation between the samples of the patient different working correlation structures investigated. Specifically first-order autoregressive, compound symmetry and independent working correlation structures were employed. Among them, first-order autoregressive working correlation assumption outperformed, give us minimum QIC value (Appendix A.2). First-order autoregressive (AR(1)) correlation structure assumes a reduction in correlation with increasing time or distance between observations. In our case, this means that the samples closer in time are highly correlated and the time gets farther and farther apart the correlation decrease.

There was no statistical test exist to assess the correctness of the working correlation structures. However, the QIC goodness of fit statistic was used to compare different working correlation assumptions. In addition, the agreement between the model-based and empirical standard errors suggests that the assumed working correlation is reasonable. The empirically corrected standard error is the one to be used, the model based is generally incorrect [Molenberghs, 2005]. The parameter estimates, empirical standard errors, and Odds ratio are presented in Table 3.11 and Table 3.12, for the outcomes ARI_1 and ARI_2 respectively.

TABLE 3.11: GEE Parameter estimates, empirical Standard errors, Odds ratio and p-value of the coefficients of predictors when the outcome is Antimicrobial resistance index one (ARI_1)

Parameter	Estimate	Std error	OR	P-value
Intercept	0.268	0.321	1.31	0.4037
treatment	-0.481	0.120	0.62	<.0001
gender	1.369	0.451	3.93	0.0024
log(time)	0.013	0.073	1.01	0.8588
log(time)*gender	-0.349	0.131	0.71	0.0078

Reference group: treatment - Fluoroquinolone; gender - Female.

Table 3.11 reports the GEE parameter estimates of the first outcome of interest, ARI_1. The output shows that a statistically significant effect of treatment on the log odds of resistance with p-value = 0.0001. The estimated odds of resistance when the treatment was aminopenicillins equal $exp(-0.481) = 0.62$ times the estimated odds of resistance when the treatment was fluoroquinolones holding other variables at specific fixed value. Regarding percent change, we can say that the odds of resistance for aminopenicillins

are 38% lower than the odds of resistance for fluoroquinolones. The result suggests that the commensal bacteria such as *E. coli* developed resistance over broad-spectrum antibiotics aminopenicillins or fluoroquinolones. Moreover, the effect of the variable log(time) depends on the gender of the patients, which is known as an interaction effect. More explicitly, we can say that for female patients, a one unit increase in log(time) lead to a change in log odds of resistance by 0.013. On the other hand, for male patients, a one unit increase in log(time) yields a change in log odds of resistance by (0.013+(-0.349)) =-0.336. Concerning the odds ratio, we can say that for female patients, the odds ratio is $exp(0.013) = 1.013$ for a one unit increase in log(time) and the odds ratio for male patients is $exp(-0.336) = 0.715$ for a one unit increase in log(time).

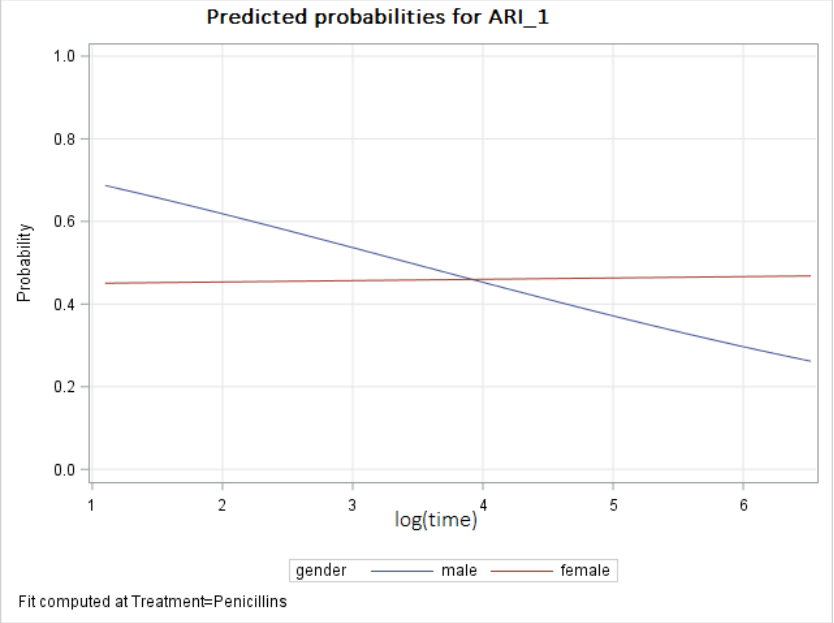


FIGURE 3.1: The plot of predicted probabilities of resistance of *E.coli* isolates over aminopenicillins and fluoroquinolones antibiotics which shows an interaction between gender and log(time).

In more detail, to explore the nature of interaction effects the predicted probability of resistances was plotted and provides an attractive visual illustration of the interaction effects. Figure 3.1 depicts that the probability of resistance of male patients is higher than female patients before 49 days(log(time)=3.9) and conversely, after 49 days female patients have a high probability of resistance than male patients.

TABLE 3.12: GEE Parameter estimates, empirical Standard errors, Odds ratio and p-value of the coefficients of predictors when the outcome is Antimicrobial resistance index two(ARI_2)

Parameter	Estimate	Std error	OR	P-value
Intercept	-0.252	0.342	0.78	0.4618
log(time)	0.039	0.083	1.04	0.6355
treatment	-0.585	0.086	0.56	<.0001
gender	0.859	0.475	2.36	0.0699
log(time)*gender	-0.300	0.128	0.74	0.0189
treatment*gender	0.549	0.132	1.73	<.0001

Reference group: treatment - fluoroquinolone; gender - female.

Likewise, GEE technique performed in the multiple logistic regression model built previously for the second outcome of interest, ARI_2. The outputs staged in Table 3.12. The presence of significant interaction terms to a fitted model drastically changes the interpretation of the coefficients of treatment, gender, and log(time). For male patients a fitted model can be written as:-

$$logit(\pi) = 0.607 - 0.261 * log(time) - 0.036 * treatment$$

For male patients, for a one unit change in log(time), the log odds of resistance decrease by -0.261. The odds ratio of treatment is 0.96, which indicates that the odds that a male patient develops resistance were 3.5% lower for patients who received aminopenicillins compared to patients who received fluoroquinolones holding other variable at specific fixed value.

Similarly, for female patients a fitted model can be written as:-

$$logit(\pi) = -0.252 + 0.039 * log(time) - 0.585 * treatment$$

However, in case of female patients for a one unit change in log(time), the log odds of resistance increase by 0.039. The odds ratio of treatment is 0.56, which means that the odds that a female patient develops resistance were 44% lower for patients who received aminopenicillins compared to patients who received fluoroquinolones keeping another variable at particular fixed value.

Furthermore, predicted probabilities were plotted to visualize the interaction between gender and log(time) at a fixed level of treatment and an interaction between gender and treatment at a fixed level of log(time). From Figure 3.2, we can see that the probability of resistance of male patients is higher than female patients approximately until 110 days(log(time)=4.7). Following 110 days onwards the probability of resistance occurs high for the female patient compared with the male patient.

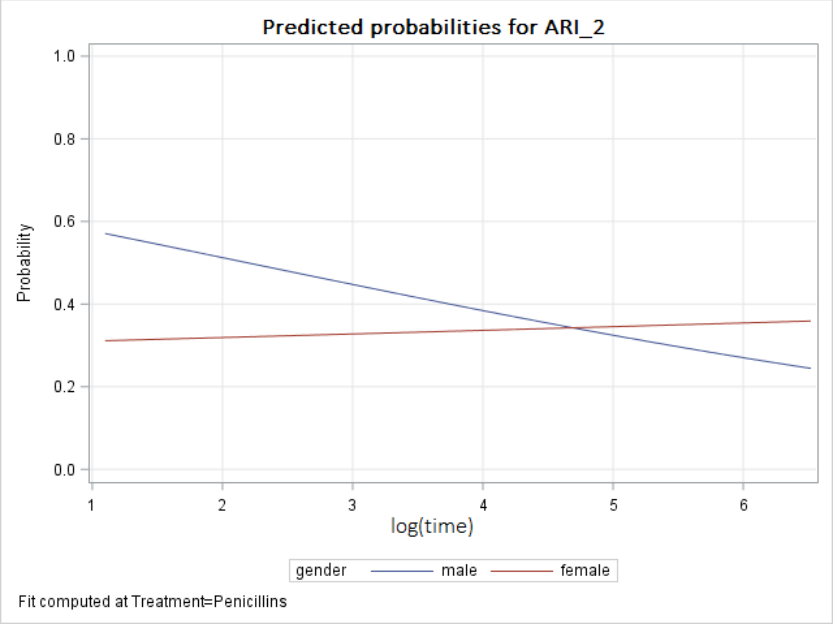


FIGURE 3.2: The plot of predicted probabilities of resistance of *E.coli* isolates over 'J01 antibacterials for systemic use' which shows an interaction between gender and log(time).

Figure 3.3 depicts that the male patient treated with aminopenicillins have a higher predicted probability of resistance than the female patient prescribed with aminopenicillins. In contrary, the male patient treated with fluoroquinolone have a lower predicted probability of resistance than the female patient administered with the same antibiotics.

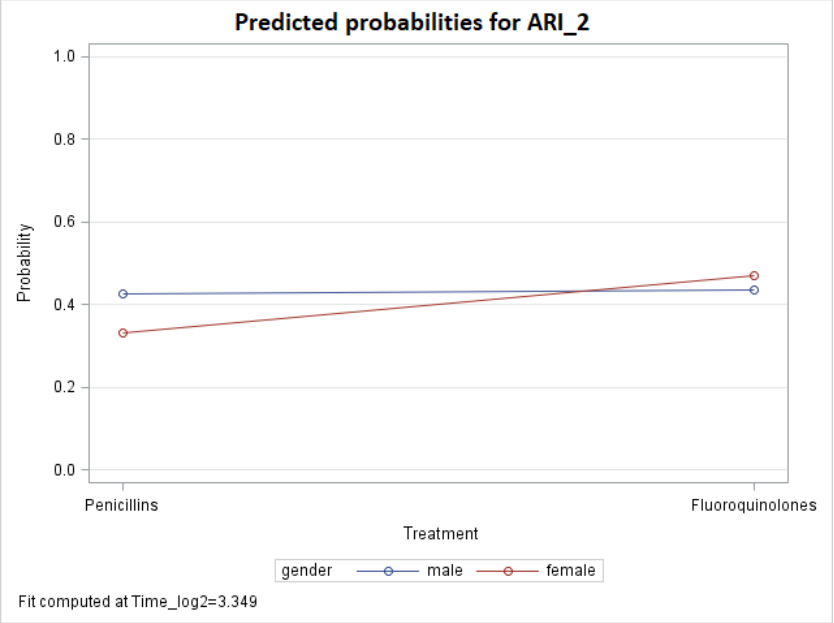


FIGURE 3.3: The plot of predicted probabilities of resistance of *E.coli* isolates over 'J01 antibacterials for systemic use' which shows an interaction between gender of the patient and treatment administrated.

Chapter 4

Discussion and Conclusion

Many studies showed that resistance of *E.coli* in the gut related to several factors. This bacterium is often resistant to the antimicrobial agents, such as aminopenicillins and fluoroquinolones usually recommended for the treatment of infection caused by E. coli [Bonfiglio et al., 2002, Rodríguez-Baño et al., 2008]. Commonly, bacterium develops resistant through mutations of target genes or horizontal transfer of genes. The lower gastrointestinal tract is the comprehensive microbial community in the human host. Bacterial biofilms are located in the human intestinal tract and are contemplated to be ideal environments for horizontal gene transfer [Huddleston, 2014].

This study aimed to assess to what extent and how long the commensal bacteria such as *E.coli* in the digestive tract(gut) affected by broad-spectrum antibiotics, such as aminopenicillins and/or fluoroquinolones. Also, whether or not this impact changed from the effect of other types of antibiotics tested in the laboratory.

The dataset used in this study obtained from a multicenter cohort study during 2010 -2012 and patient prescriptions from national reimbursement data (collected by the inter mutualistic agency – IMA, Brussels). It is proper to summarize a group of independent observations by the number of observations in the group that represent one of the two outcomes. Each sample taken tested with different antimicrobial in the laboratory, and the test result grouped into either resistance or susceptible. Antimicrobial resistance index used as an outcome variable, which is the proportion of resistance to multiple antimicrobial tested for each sample.

To reach the aim of the study antimicrobial resistance index was computed in two different contexts. The first is only antibiotics fluoroquinolones and aminopenicillins tested in the laboratory and the second is the whole antibiotics 'J01' tested in the laboratory. Statistical methods have been used to analyze two response variables distinctly together with given predictor variables. A binomial distribution assumed for the response variables. Since more than one samples taken per patient so, there could be a correlation between the samples of the patient. To account the correlations between samples within the patient generalized estimating equation with first-order autoregressive working correlation structure was adopted.

In the GEE analysis of the first outcome, we noticed that the treatment has a significant effect on the log odds of resistance. According to EARS-net report in 2015, among *E.coli* isolates reported more than half were resistant to at least one antimicrobial group under surveillance. Inside their report resistance to aminopenicillin and fluoroquinolones has most often described, both in single resistance and in combination with other antimicrobials ECDC [2017]. Similarly, in our study, the commensal bacteria such as *E. coli* retrieved from the digestive tract(gut) developed resistance to aminopenicillins and fluoroquinolones antibiotics. Besides, we detected that an interaction effect between gender of the patient and log(time), number of days between prescription date and sampling date, on the log odds of resistance.

The result of the second outcome of interest, ARI_2, showed that the effect of gender on the log odds of resistance has interacted with log(time). It means that the impact of the number of days between prescription date and sampling date on the log odds of resistance depend on the gender the patient. In more detail as the number of days increase the probability of resistance decrease for male patients and on the contrary slight increase for the female patient. It might be due to female have a constantly high antimicrobial resistance index due to over treatment in the past. A meta-analysis by Schroder et al., showed that females receive a high number of prescription than male patients [Schröder et al., 2016]. They also suggested that nowadays the evidence on infectious-disease epidemiology by gender cannot fully clarify this substantial difference. We recommend carrying out further analysis to reason out why such differences have occurred and their impact on the development of resistance.

Furthermore, we also also found that a statistically significant interaction effect between treatment given to the patient and gender of the patient on the log odds of resistance.

It might be due to urinary tract infection treatments were specifically given to female patients, and rarely given to male patients. An earlier study done by Sahuquillo-Arce et al., confirmed the presence of a definite difference in antimicrobial resistance between *E.coli* isolates from male and female patients [Sahuquillo-Arce et al., 2011]. In conclusion, based on this study the *E.coli* had developed resistance to broad-spectrum antibiotics namely aminopenicillins and fluoroquinolones but it depend on the gender of the patient.

Appendix A

Tables

TABLE A.1: A table shows the total number of aminopenicillins and fluoroquinolones antibiotics prescribed by administration route.

Frequency Percent

Treatment	Route of administration		
	Oral	Parenteral	Total
Fluoroquinolones	72 30.51	3 1.27	75 31.78
Penicillins	122 51.69	39 16.53	161 68.22
Total	194 82.2	42 17.8	236 100

TABLE A.2: A goodness of fit statistic (QIC) of the two models fitted using GEEs techniques with different working correlation structure.

	Model one	Model two
Working Correlation	QIC	QIC
First order Autoregressive	175	81
Compound symmetry	178	83
Independent	201.79	98

NB: Model one: Outcome of antimicrobial resistaance index one (ARI_1)

Model two: Outcome of antimicrobial resistance index two (ARI_2).

TABLE A.3: A table shows total number of *E.coli* isolates tested in each of the laboratories.

lab_id	Frequency	Percent	Cumulative Frequency
105	6	2.54	6
108	2	0.85	8
109	222	94.07	230
110	5	2.12	235
111	1	0.42	236

Appendix B

Figures

[1]

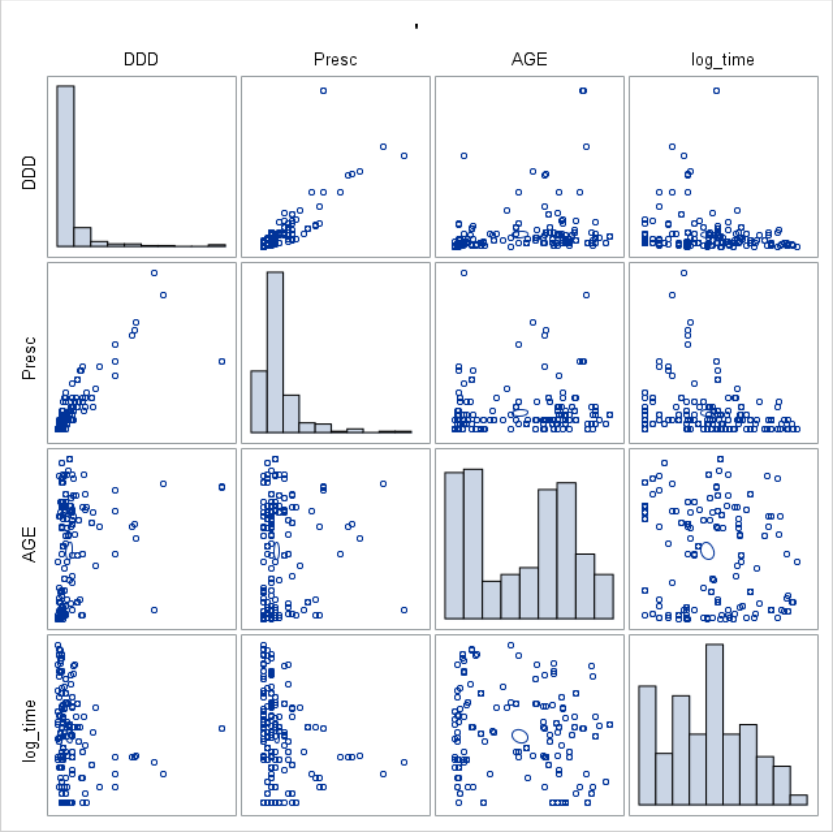


FIGURE B.1: Scatter plot and histogram of continuous predictor variables included in the study.

[1] **NB**: DDD: defined daily dose and Presc: Number of prescription before sample.

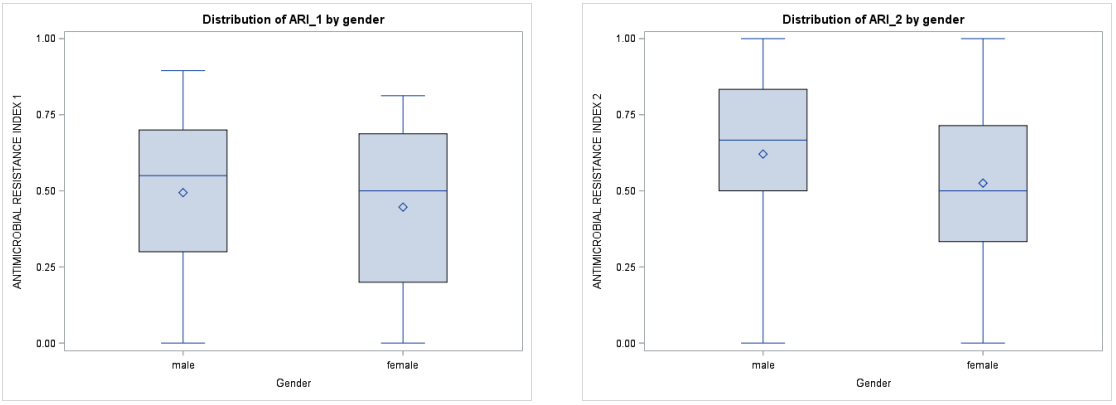


FIGURE B.2: Box-plot shows the distribution of antimicrobial resistance indexes(ARI) by the gender of the patient. Left: (ARI_1), Right: ARI_2.

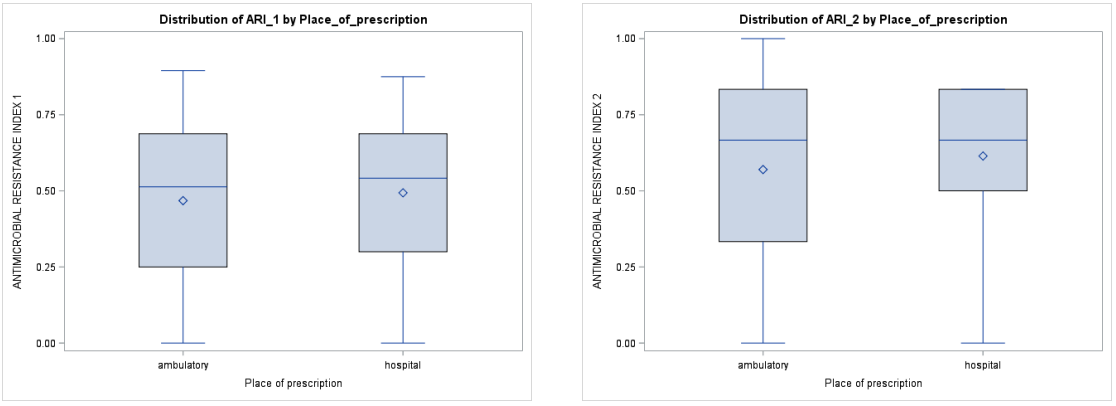


FIGURE B.3: Box-plot shows the distribution of antimicrobial resistance indexes(ARI) by the place of prescription. Left: (ARI_1), Right: ARI_2.

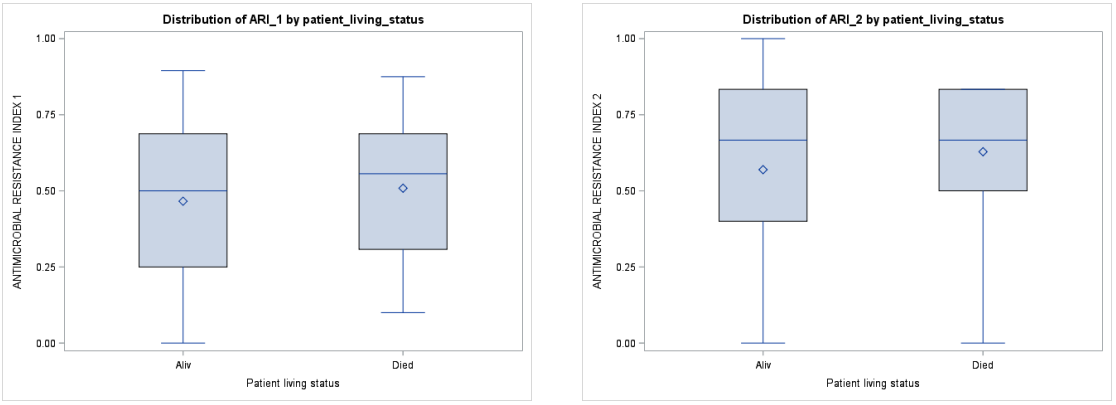


FIGURE B.4: Box-plot shows the distribution of antimicrobial resistance indexes(ARI) by living status of the patient up to the end of the study year. Left: (ARI_1), Right: ARI_2.

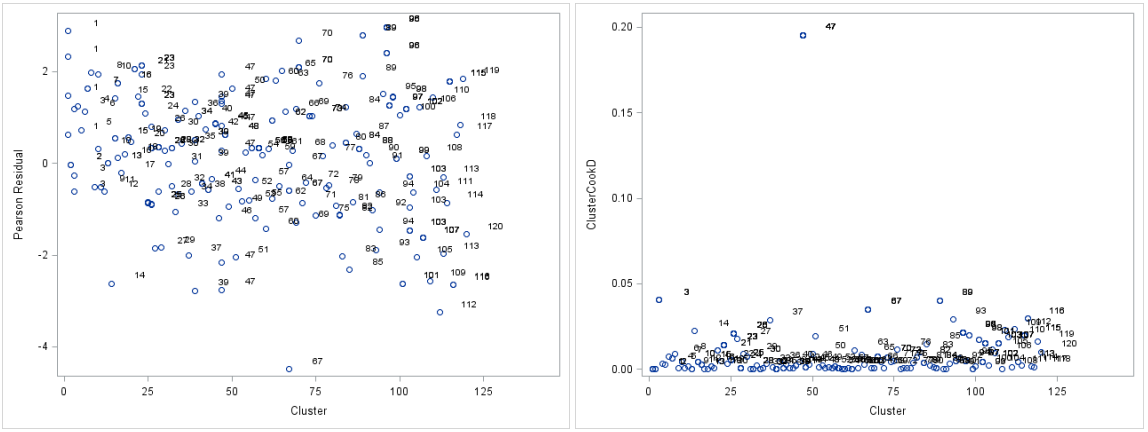


FIGURE B.5: Diagnostic plots of the first outcome of interest (ARI_1) Left: Scatter Plot of Pearson residual Right: Scatter Plot of Cook-distance at cluster level

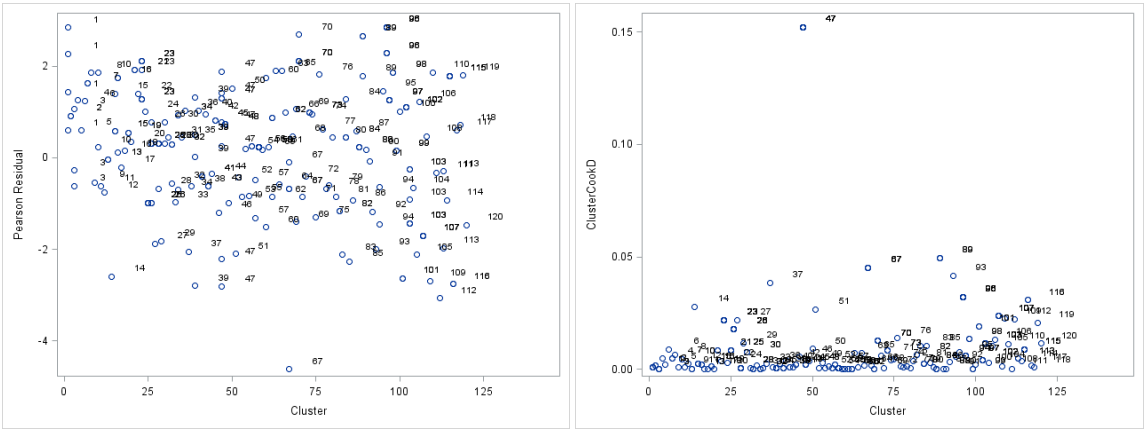


FIGURE B.6: Diagnostic plots of the second outcome of interest (ARI_2)) Left: Scatter Plot of Pearson residual Right: Scatter Plot of Cook-distance at cluster level

Bibliography

Agresti, A. (1996). *An introduction to categorical data analysis*, volume 135. Wiley New York.

Bailey, J. K., Pinyon, J. L., Anantham, S., and Hall, R. M. (2010). Commensal escherichia coli of healthy humans: a reservoir for antibiotic-resistance determinants. *Journal of medical microbiology*, 59(11):1331–1339.

Bonfiglio, G., Simpore, J., Pignatelli, S., Musumeci, S., and Solinas, M. L. (2002). Epidemiology of bacterial resistance in gastro-intestinal pathogens in a tropical area. *International journal of antimicrobial agents*, 20(5):387–389.

Czepiel, S. A. (2002). Maximum likelihood estimation of logistic regression models: theory and implementation. *Available at czep. net/stat/mlelr. pdf*.

Dellit, T. H., Owens, R. C., McGowan, J. E., Gerding, D. N., Weinstein, R. A., Burke, J. P., Huskins, W. C., Paterson, D. L., Fishman, N. O., Carpenter, C. F., et al. (2007). Infectious diseases society of america and the society for healthcare epidemiology of america guidelines for developing an institutional program to enhance antimicrobial stewardship. *Clinical infectious diseases*, 44(2):159–177.

ECDC (2017). Antimicrobial resistance surveillance in europe 2015. *Annual Report of the European Antimicrobial Resistance Surveillance Network (EARS-Net), Stockholm*.

Etebu, E. and Arikekpar, I. (2016). Antibiotics: classification and mechanisms of action with emphasis on molecular perspectives. *Int J Appl Microbiol Biotechnol Res*, 4:90–101.

Fair, R. J. and Tor, Y. (2014). Antibiotics and bacterial resistance in the 21st century. *Perspectives in medicinal chemistry*, 6:PMC–S14459.

Finch, R. G., Greenwood, D., Whitley, R. J., and Norrby, S. R. (2010). *Antibiotic and chemotherapy e-book*. Elsevier Health Sciences.

Hilbe, J. (2009). *Logistic Regression Models*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press.

Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.

Huddleston, J. R. (2014). Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. *Infection and drug resistance*, 7:167.

Hutchinson, J. M., Patrick, D. M., Marra, F., Ng, H., Bowie, W. R., Heule, L., Muscat, M., and Monnet, D. L. (2004). Measurement of antibiotic consumption: A practical guide to the use of the anatomical therapeutic chemical classification and defined daily dose system methodology in canada. *Canadian Journal of Infectious Diseases and Medical Microbiology*, 15(1):29–35.

Kadane, J. B. and Lazar, N. A. (2004). Methods and criteria for model selection. *Journal of the American statistical Association*, 99(465):279–290.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Lin, J., Nishino, K., Roberts, M. C., Tolmasky, M., Aminov, R. I., and Zhang, L. (2015). Mechanisms of antibiotic resistance. *Frontiers in microbiology*, 6:34.

Llor, C. and Bjerrum, L. (2014). Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. *Therapeutic advances in drug safety*, 5(6):229–241.

MacDougall, C., Powell, J. P., Johnson, C. K., Edmond, M. B., and Polk, R. E. (2005). Hospital and community fluoroquinolone use and resistance in staphylococcus aureus and escherichia coli in 17 us hospitals. *Clinical infectious diseases*, 41(4):435–440.

Martin, S. J., Micek, S. T., and Wood, G. C. (2010). Antimicrobial resistance: consideration as an adverse drug event. *Critical care medicine*, 38:S155–S161.

McGahan, C. E. (2017). Using a population average model to investigate the success of a customer retention strategy. *British Columbia Cancer Agency*.

Molenberghs, G.and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer-Verlag New York.

Munita, J. M. and Arias, C. A. (2016). Mechanisms of antibiotic resistance. *Microbiology spectrum*, 4(2).

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996). *Applied linear statistical models*, volume 4. Irwin Chicago.

Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics*, 57.

Rodríguez-Baño, J., Alcalá, J. C., Cisneros, J. M., Grill, F., Oliver, A., Horcajada, J. P., Tórtola, T., Mirelis, B., Navarro, G., Cuenca, M., et al. (2008). Community infections caused by extended-spectrum β-lactamase–producing escherichia coli. *Archives of internal medicine*, 168(17):1897–1902.

Sahuquillo-Arce, J. M., Selva, M., Perpiñán, H., Gobernado, M., Armero, C., López-Quílez, A., González, F., and Vanaclocha, H. (2011). Antimicrobial resistance in more than 100,000 escherichia coli isolates according to culture site and patient age, gender, and location. *Antimicrobial agents and chemotherapy*, 55(3):1222–1228.

Schröder, W., Sommer, H., Gladstone, B. P., Foschi, F., Hellman, J., Evengard, B., and Tacconelli, E. (2016). Gender differences in antibiotic prescribing in the community: a systematic review and meta-analysis. *Journal of Antimicrobial Chemotherapy*, 71(7):1800–1806.

Thomas, S. and Karl, M. (2017). European antimicrobial resistance surveillance network (ears-net belgium). *Unit Healthcare associated infections and Antimicrobial resistance Brussels, Belgium*.

WHO et al. (2014). *Antimicrobial resistance: global report on surveillance*. World Health Organization.

Wright, P. M., Seiple, I. B., and Myers, A. G. (2014). The evolving role of chemical synthesis in antibacterial drug discovery. *Angewandte Chemie International Edition*, 53(34):8840–8869.

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Persistence of antimicrobial resistance in E. coli after exposure to penicillins or fluoroquinolones

Richting: **Master of Statistics-Biostatistics**
Jaar: **2018**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Haile, Yilikal Tesfaye

Datum: **15/06/2018**