Combatting Antimicrobial Resistance through a Data-Driven Approach to Optimize Antibiotic Use and Improve Patient Outcomes: Protocol for a mixed approach study

Jonathan Mayito, Ronald Galiwango, Conrad Tumwine, Suzan Nakasendwa, Mackline Hope, Andrew Kambugu, Elly Nuwamanya, Reuben Kiggundu, Flavia Dhikusooka, Vivian Twemanye, Andrew Kambugu, Francis Kakooza

Submitted to: JMIR Research Protocols
on: March 07, 2024

Table of Contents

Combatting Antimicrobial Resistance through a Data-Driven Approach to Optimize Antibiotic Use and Improve Patient Outcomes: Protocol for a mixed approach study

Jonathan Mayito[1*] MBChB, MM, PhD; Ronald Galiwango[1, 2*] BSc, MPhil, PhD; Conrad Tumwine[1*] MBChB; Suzan Nakasendwa[1*] BScStatistics, MScBiostatistics; Mackline Hope[1*] MBChB, MPH, DTMH; Andrew Kambugu[1*] MBChB, MM; Elly Nuwamanya[1*] BSc, MScEcon; Reuben Kiggundu[1*] MBChB, MScPhysiology, MPH; Flavia Dhikusooka[1*] BSc; Vivian Twemanye[1*] BPharm, MPH; Andrew Kambugu[1*] MBChB, MM; Francis Kakooza[1*] PhD

[1]Infectious Diseases Institute Kampala UG
[2]The African center of Excellence in Bioinformatics and Data Intensive Sciences The Infectious Diseases Institute College of Health Sciences, Makerere University, Kampala, Uganda. Kampala UG
[*]these authors contributed equally

Corresponding Author:
Conrad Tumwine MBChB
Infectious Diseases Institute
IDI-McKinnell Knowledge Centre, P.O. Box 22418, Makerere University, Kampala, Uganda
Kampala
UG

Abstract

Background: It is projected that drug resistant infections will lead to 10 million deaths annually by 2050, if left unabated. Despite this threat, surveillance data from resource limited settings is scarce and often lacks antimicrobial resistance (AMR)-related clinical outcomes and economic burden. We aim to build an AMR and antimicrobial use (AMU) data warehouse, describe the trends of resistance and antibiotic use, and determine the economic burden of AMR in Uganda and develop a machine learning algorithm for predicting AMR-related clinical outcomes.

Objective: The overall objective of the study is to use data-driven approaches to optimize antibiotic use and combat antimicrobial resistant infections in Uganda. We aim to 1)To build a dynamic AMR and Antimicrobial Use and Consumption (AMUC) Data Warehouse to support research in AMR and AMUC to inform AMR related interventions and public health policy, (2) To evaluate the trends in AMR and antibiotic use using annual antibiotic and point prevalence survey data collected at nine regional referral hospitals over a five-year period, (3) To develop a Machine Learning model for predicting the clinical outcomes of patients with bacterial infectious syndromes due to drug resistant pathogens and (4) To estimate the annual economic burden of AMR in Uganda using the cost-of-illness approach.

Methods: We will conduct a data curation, machine learning based modelling, and cost of illness analysis study using AMR and AMU data abstracted from procurement, human resources, and clinical records of patients with bacterial infectious syndromes at nine regional referral hospitals in Uganda, collected between 2018 and 2026 We will employ data curation procedures, FLAIR (Findable, Linkable, Accessible, Interactable and Repeatable) principles and role-based access controls (RBAC) to build a robust and dynamic AMR and AMU data warehouse. We will also apply machine learning algorithms to model AMR-related clinical outcomes, advanced statistical analysis to study AMR and AMU trends, and cost of illness analysis to determine AMR-related economic burden.

Results: Once implemented, the expected results from the study will include a robust and dynamic data warehouse, AMR and AMU trends, a machine learning model for predicting AMR-related clinical outcomes (hospital lengthen of stay, time to clinical improvement or negative cultures, medical complications, mortality, and disability), and the cost per AMR case to describe the AMR-related economic burden.

Conclusions: The data warehouse will promote access to rich and interlinked AMR and AMU datasets to answer AMR program and research questions using a wide evidence base. The AMR-related clinical outcomes model and cost data will facilitate improvement in the clinical management of AMR patients and guide resource allocation to support AMR surveillance and

interventions. Clinical Trial: Infectious Diseases Institute Research Ethic Committee (IDI-REC-2023-67:) & Uganda National Council for Science and Technology (UNCST - HS3690ES).

(JMIR Preprints 07/03/2024:58116)
DOI: https://doi.org/10.2196/preprints.58116

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Combatting Antimicrobial Resistance through a Data-Driven Approach to Optimize Antibiotic Use and Improve Patient Outcomes: Protocol for a mixed approach study

Jonathan Mayito[1], Conrad Tumwine[1], Andrew Kambugu[1], Suzan Nakasendwa[1], Ronald Galiwango[1,2], Elly Nuwamanya[1], Hope Mackline[1], Reuben Kiggundu[1], Dathan M. Byonanebye[1], Francis Kakooza[1]

Author affiliation

1. Infectious Diseases Institute, Makerere University College of Health Sciences, Kampala, Uganda.

2. The African center of Excellence in Bioinformatics and Data Intensive Sciences, The Infectious Diseases Institute, College of Health Sciences, Makerere University, Kampala, Uganda.

Abstract

Background

It is projected that drug resistant infections will lead to 10 million deaths annually by 2050, if left unabated. Despite this threat, surveillance data from resource limited settings is scarce and often lacks antimicrobial resistance (AMR)-related clinical outcomes and economic burden. We aim to build an AMR and antimicrobial use (AMU) data warehouse, describe the trends of resistance and antibiotic use, and determine the economic burden of AMR in Uganda and develop a machine learning algorithm for predicting AMR-related clinical outcomes.

Objectives.

The overall objective of the study is to use data-driven approaches to optimize antibiotic use and combat antimicrobial resistant infections in Uganda. We aim to 1)To build a dynamic AMR and Antimicrobial Use and Consumption (AMUC) Data Warehouse to support research in AMR and AMUC to inform AMR related interventions and public health policy, (2) To evaluate the trends in AMR and antibiotic use using annual antibiotic and point prevalence survey data collected at nine regional referral hospitals over a five-year period, (3) To develop a Machine Learning model for predicting the clinical outcomes of patients with bacterial infectious syndromes due to drug resistant pathogens and (4) To estimate the annual economic burden of AMR in Uganda using the cost-of-illness approach.

Methods

We will conduct a data curation, machine learning based modelling, and cost of illness analysis study using AMR and AMU data abstracted from procurement, human resources, and clinical records of patients with bacterial infectious syndromes at nine regional referral hospitals in Uganda, collected between 2018 and 2026 We will employ data curation procedures, FLAIR (Findable, Linkable, Accessible, Interactable and Repeatable) principles and role-based access controls (RBAC) to build a robust and dynamic AMR and AMU data warehouse. We will also apply machine learning algorithms to model AMR-related clinical outcomes, advanced statistical analysis to study AMR and AMU trends, and cost of illness analysis to determine AMR-related economic burden.

Results

Once implemented, the expected results from the study will include a robust and dynamic data warehouse, AMR and AMU trends, a machine learning model for predicting AMR-related clinical outcomes (hospital lengthen of stay, time to clinical improvement or negative cultures, medical complications, mortality, and disability), and the cost per AMR case to describe the AMR-related economic burden.

Conclusion

The data warehouse will promote access to rich and interlinked AMR and AMU datasets to answer AMR program and research questions using a wide evidence base. The AMR-related clinical outcomes model and cost data will facilitate improvement in the clinical management of AMR patients and guide resource allocation to support AMR surveillance and interventions.

Trial Registration: Infectious Diseases Institute Research Ethic Committee (IDI-REC-2023-67:) & Uganda National Council for Science and Technology (UNCST - HS3690ES).

Keywords: Antimicrobial Resistance; AMR-Database; AMR; Machine-Learning; Antimicrobial use; Artificial Intelligence; CAMO-Net

Introduction

Antimicrobial resistance is among the greatest threats to global health and is predicted to become the leading cause of death by 2050 [1]. Mortality due to AMR will rise from 700, 000 deaths to 10 million deaths annually if nothing is done to halt the current trends [2]. Indeed, in 2019, only eight years from when these projections were made, 1.27 million deaths were attributed to AMR, higher than those due to Human Immuno-deficiency Virus (HIV), malaria, and tuberculosis [3]. Developing countries disproportionately bear the AMR burden. For instance, the all-cause mortality in the Murry et al review was highest in developing countries: 27.3 per 100, 000 in Western Sub-Saharan Africa (SSA) compared to 6.5 per 100,000 in Australasia [3]. Despite this worrying burden in sub-Saharan Africa, AMR data remains scanty, and the systems to monitor and generate AMR data are underdeveloped. The region had a low Joint External Evaluation score of 53% and many of the countries in the region lacked national action plans (NAP) for AMR, alluding to AMR not being among their health priorities [4].

AMR is a natural phenomenon, and tends to arise from enzymatic degradation, alterations in antimicrobial targets, or change in membrane permeability to the antimicrobials [5]. It is accelerated by the misuse and overuse of antimicrobials, poor infection prevention and control, limited access to quality affordable medicines, vaccines, and diagnostics, lack of awareness, and poor enforcement of prescription regulations [1]. Infections due to resistant microbes are more difficult to treat, spread more, and are of higher severity, resulting in increased morbidity, mortality, and cost of healthcare [1]. Antimicrobial misuse and overuse are the most critical drivers of AMR. For example, a review of Slovenia's surveillance data showed that the prevalence of invasive *Streptococcus pneumoniae* resistance to penicillin decreased by 47.1% following a 32.8 % decline in AUC [6]. In the same study, high consumption of clarithromycin resulted in the selection and predominance of macrolide-resistant *Streptococcus pneumonia*. Therefore, given the little global effort to develop new antimicrobials [7], the control of misuse and overuse presents the best chance to tackle AMR, more so in resource-limited settings.

AMR is not without consequences to the individual and health systems. Treatment of AMR infections necessitates the use of more expensive and potentially toxic medications, results in longer hospital stay and ultimately into higher cost of treatment[8]. In addition, resistant infections pose a risk to specialized medical procedures—such as organ transplants, surgical operations, and cancer chemotherapy — denying many these critically needed services [8]. Mortality from infections is also

increasing; AMR was the leading cause of death in 2019, surpassing HIV and tuberculosis [3]. Furthermore, the Disability Adjusted Life Years (DALYs), are also increasing, indicating a reduction in life expectancy and quality of life [9]. Despite the importance of these parameters in characterizing the AMR burden, they are not routinely available, hindering comprehensive efforts to control the spread and effects of AMR.

The proposed project therefore aims to close four important gaps in AMR surveillance and interventions: 1) Improve access to AMR and AMUC data; 2) describe the AMR and antibiotic prescribing patterns, changes in the patterns over time, and associated factors; 3) determine the economic burden of excessive antibiotic use and the effect on hospital budgets; and 4) develop a machine learning model to predict the clinical outcomes of bacterial infectious syndromes caused by resistant pathogens. We hypothesize that this study will contribute to AMR/AMUC surveillance, identification of priority antibiotics to target for stewardship interventions, determining of the cost implications and identifying characteristics of individuals more likely to have poor clinical outcomes related to AMR so as to guide clinical decision and health system planning in the allocation of scarce resources.

The overall objective of the study is to use data-driven approaches to optimize antibiotic use and combat antimicrobial resistant infections in Uganda.

The specific study objectives will include;

1. To build a dynamic AMR and Antimicrobial Use and Consumption (AMUC) Data Warehouse to support research in AMR and AMUC to inform AMR related interventions and public health policy.

2. To evaluate the trends in AMR and antibiotic use using annual antibiotic and point prevalence survey data collected at nine regional referral hospitals over a five-year period.

3. To develop a Machine Learning model for predicting the clinical outcomes of patients with bacterial infectious syndromes due to drug resistant pathogens.

4. To estimate the annual economic burden of AMR in Uganda using the cost-of-illness approach.

Study methods

Study design

The study will employ different study designs for each of the four specific objectives as detailed below;

Objective one will be a data curation project aimed at building an AMR/AMUC data warehouse from prior and future AMUC surveys and routine AMR surveillance data from patients with bacterial infectious syndromes.

Objective two will be a retrospective study using annual antibiotic surveys and point prevalence survey (PPS) data to determine the trends in antibiotic use and AMR using data accrued over a five-year period.

Objective three will be a modeling study in which a machine-learning model will be developed to predict the clinical outcomes of patients with bacterial infectious syndromes due to resistant pathogens.

Objective four will be a prevalence-based cost of illness (COI) descriptive study to assess the economic burden of AMR in Uganda by determining the average cost per AMR case.

Study settings

This study will be based on data already accrued under the Fleming Fund project or abstracted from clinical and other records at the Fleming Fund project supported surveillance sites during the study period. The Fleming Fund Country Grant project (2018 to 2026), is a health system strengthening project to improve AMR surveillance at nine Regional Referral Hospitals (RRH) in Uganda. The surveillance sites include Arua RRH, Gulu RRH, Lira RRH, Soroti RRH, Mbale RRH, Jinja RRH, Masaka RRH, Mbarara RRH, and Kabale RRH. The RRH, which are spread across the different regions of the country, are the first level of specialized healthcare in Uganda, with specialist health workers and services.

Study population

The study will use records of patients with bacterial infectious syndromes attending the surveillance sites for health care.

Inclusion criteria

All medical records of patients with bacterial infectious syndromes attending the major hospital wards including the medical, surgical, pediatric, gynecology, and maternity wards, and the out-patients department are eligible to include in the study. Further, records related to medicines and supplies procurement as well as human resource and utility bills will also be targeted by the study for the cost of illness evaluation.

Exclusion criteria

For the AMR data, records without the results for antimicrobial susceptibility testing (AST) results will be excluded from the study.

Study procedures

Case ascertainment

The patients with bacterial infectious syndromes will be identified from the diagnoses made by the attending clinicians and recorded in the patients' clinical records. The resistant infections are determined from the AST results from the cultured samples showing resistance to at least one antibiotic.

Data sources

Data will be abstracted from the procurement and human resources records, and the clinical records of patients with bacterial infectious syndromes attending Arua RRH, Gulu RRH, Lira RRH, Soroti RRH, Mbale RRH, Jinja RRH, Masaka RRH, Mbarara RRH, and Kabale RRH.

Data elements collected

The treatment-related variables collected include referral status, age, gender, occupation, ward, residence, prior antibiotic treatment, date of admission, date of sample collection, sample type, diagnosis, and AST results. The AMUC variables collected include ward, number of prescribed antibiotics, number of prescriptions, number of injectable antibiotics prescription, prescriptions by generic name, prescriptions according to guidelines, prescriptions with appropriate diagnosis, cultures requested, prescriptions based on AST, missed doses, patients with missed doses, referrals, days spent on the ward, and no. of hospitalizations in last 90 days.

The clinical parameters to be collected will include temperature, blood pressure, level of antibiotic resistance, antibiotic type, diagnosis, prior antibiotic exposure, site of infection, duration of

symptoms, timing of effective antimicrobial treatment, and medical comorbidities, while the clinical outcomes will include duration of hospital stay, time to clinical improvement, time to negative cultures, mortality, medical complications, and disability), and the cost of treatment.

The cost of illness variables will include direct medical costs (personnel, medical supplies, drugs, laboratory tests and patient out-of-pocket costs), direct non-medical costs (recurrent expenditures - such as utility bills) and capital expenditures (such as expenditures on hospital/ health facility infrastructure), patient transportation, and upkeep while seeking medical care. Indirect costs will include productivity losses due to illness-related absenteeism, reduced work hours, disability, premature mortality, and informal care provided by family members or friends.

Data abstraction procedure

Antimicrobial resistance data is collected during the routine care of patients with bacterial infectious syndromes at the surveillance sites. The process includes the identification of patients requiring sampling, drawing of the sample using appropriate techniques, transportation of the sample to the laboratory, sample integrity assessment, sample processing, recording of results on paper and electronic data systems, and reporting of results for clinical management of the patients. On the other hand, AMUC data is collected through quarterly point prevalence surveys (PPS) and annual antibiotic surveys (AAS). During PPS, data is abstracted from records of all patients on the wards for at least 24 hours while for the AAS data is abstracted from 100 randomly selected records of patients seen at a particular ward in the last year. During PPS and annual antibiotic surveys, data is collected from six hospital units including the outpatient department; and medical, surgical, gynecology, pediatric, and maternity wards. The clinical outcome and cost of illness data will also be abstracted from the clinical, procurement and human resources records along-side the PPS data over three quarters. The clinical outcome data will be collected from the five hospital units including the medical, surgical, gynecology, pediatric and maternity wards.

The laboratory request form, laboratory results register, and the WHONET are used to collect the AMR data while AMUC data is collected using the World Health Organization (WHO) PPS tool and a standardized excel tool based on the WHO/INRUD drug use indicator for the AAS data. The PPS tool is built in open data kit (ODK), an android App with offline capabilities installed on mobile smart devices including mobile phones and tablet computers, used for collecting, managing, and using data. The abstraction tool was pretested to evaluate the validation rules in ODK and to ensure the completeness, reliability, and validity of the tools. The AAS data is collected annually over a

period of two weeks for each facility while the PPS data is collected quarterly over a period of two days for each facility. The clinical outcome and cost of illness data will be collected quarterly using data abstraction forms built in REDCap [10].

Hospital pharmacists who are supervised by the senior pharmacists and doctors abstract the data after being trained on the data collection tools and the approach to collecting antimicrobial use data. The supervisors review each completed form immediately after collection and the finalized form is submitted to the server, after which it cannot be edited further. Appropriate use is assessed using the current Uganda Clinical Guidelines (UCG).

Sample size

Objective one

All patients with bacterial infectious syndromes who access the surveillance sites for routine medical care are the primary source of the AMR data. The project has data on 12,366 patients over a last four years. We have also collected AMUC data on 13,500 patients over an eight-year period. We will also supplement these data with PPS data on 18,000 patients from AAS, giving a total of 31,500 patients' records. As more data is gathered using the aforementioned data-gathering techniques, the sample size will continue to increase for the periodic assessments.

Objective two

The main outcome is the trends in the prevalence of antibiotic prescriptions. Using the formula for sample size estimation for a proportion; sample size = p(1-p)*(z/e)^2 * n/[(1+(n-1))*r]. Where z is the z-score corresponding to the desired confidence level. In this case z = 1.96 for 95% confidence interval. p is the estimated prevalence of antibiotic use, taken at 74% as prevalence determined by Kiggundu et al in 13 hospitals in Uganda [11]. e as the margin of error, expressed as a proportion, assumed 0.03. n as the number of clusters, in this case the 9 surveillance sites and r as the estimated intra-cluster correlation coefficient. The prevalence shown to be similar across our surveillance, hence used r = 0.9 [11]. **Sample size for each round of survey** = 0.74*(1-0.74)*(1.96/0.03)^2 * 9/((1+(9-1))*0.9) = 903.

Medical records of 100 patients from each of the major units including medical, surgical, pediatric, gynecology, and obstetrics/maternity ward, and the out-patients department are selected for the annual antibiotic survey. Therefore, 600 medical records are selected at each of the 9 surveillance sites during each survey thus, 27,000 (600*9*5) records in five years. On the other hand, on average

20 to 30 records are included in the PPS from each ward, and 100 to 150 from the five wards considered for each of the 9 hospitals. Therefore 9000 (9*5*20*10) to 13,500 (9*5*30*10) records have been used for the last 10 PPS carried out to date. A total of at least 40,500 (27,000 + 13,500) records will provide data for this evaluation. **Therefore, the accrued data is sufficiently powered to estimate the prevalence and its trend over the evaluation period.**

Objective three

Generally, there is no fixed sample size for machine learning algorithms but the more data, the better the accuracy of the model in predicting the outcome. However, for prediction models the rule of thumb for the sample size is to have at least 10 events for each predictor variable [12, 13]. Therefore, the 18 variables that will be abstracted in section 2.5.7 would require at least **18x10 = 180** records. However, the study will utilize all available data of patients with bacterial infectious syndromes due to resistant and susceptible bacterial infections accrued through quarterly data abstraction. Seventy percent (70%) of the data will be used to train while 30% will be used to test the machine learning algorithm in predicting the clinical outcomes of the patients with bacterial infectious syndromes.

Objective four

Cost of illness or economic burden evaluations unlike cost effectiveness studies are not typically based on a particular sample size [14]. The study will therefore utilize data of all patients with bacterial infectious syndromes due to resistant and those with susceptible bacterial infections and other related costs accrued through the quarterly data abstraction for this AMR economic burden evaluation.

Statistical analysis

Objective one

Expected outcome

The expected outcome is a refined and dynamic AMR and AMUC Data Warehouse, which is linkable to other databases to enable collaborative research and programing for decision-making, policy formulation, and quality improvement projects.

Building the Data Warehouse

Building the Data Warehouse will comprise all data curation steps including cleaning, merging, cataloging, and integration. The data variables will be aligned to national and international AMR and

AMUC indicators for easy alignment to contemporary literature. The curated data will be uploaded to a secure repository, which will be continuously updated as more data is accrued. The data repository will be housed in the African Center of Excellence in Bioinformatics & Data-intensive Sciences High-Performance Scientific Computing Infrastructure (ACE-HPC) information technology infrastructure at IDI. The HPC is supported by a 30KVA inverter system that provides extra uptime in the event of a power outage and a monitoring system that monitors HPC systems for power surges and internet outage. Access to the HPC system is via a password protected user account that is associated to an active email address with a mandatory training session for all first-time users of the HPC. Access to the servers will be via Secure Shell protocol (SSH) to respective login nodes with access to compute nodes via the job scheduler (Slurm). User data directories and shared data directories will be backed up to enable recovered in the event of data loss.

The warehouse will adhere to the FAIR principles: 1) Findable – where the data is easy to find by humans and computers/machines metadata standards and tags based on a defined criteria, 2) Accessible: found data would be accessible through appropriate authentication and authorization, 3) Interoperable: the accessed data would be interactable with other data-sets through different applications and workflows for analysis, storage, and processing, and 4) Reusable – the data will be well indexed so that it can be replicated or combined with other data sets to interrogate wide research, surveillance and stewardship questions.

Access to the Data Warehouse will be managed by the Role-based access control (RBAC) for the warehouse developers, where system administrators will assign user roles and manage the access for each role. For the researchers and program investigators, the process to access the data in the warehouse will include: submitting data request form and a concept on the intended use. The data request form will specify the type of data and the specific variables required. A scientific committee will review the concept and will recommend whether access should be granted. Before the data is availed, the researcher or investigator will sign a data use agreement defining the terms under which the data will be used.

Objective two

The objective two outcome variables will be derived as indicated in the table below.

Indicator	**Indicator definition**	**Disaggregation**
Antimicrobial Use:		

Indicator	Indicator definition	Disaggregation
(i) Measuring the impact on selected drug use indicators		
Proportion of prescriptions with at least one antibiotic	N: Total number of antibiotic prescriptions D: Total number of patients included in the sample	Gender of the patients Age of the patients
Proportion of prescriptions with an injectable antibiotic	N: Total number of antibiotic prescriptions with an injectable antibiotic D: Total number of antibiotic prescriptions	Gender of the patients Age of the patients
Proportion of antibiotic prescriptions with a diagnosis that does not warrant antibiotic	N: number of antibiotic prescriptions with a diagnosis that doesn't warrant an antibiotic D: total number of antibiotic prescriptions	Gender of the patients Age of the patients
Proportion of antibiotic prescriptions for *upper respiratory infections	N: total number of antibiotic prescriptions for upper respiratory tract infections D: total number patients diagnosed with a URTI	Gender of the patients Age of the patients
Proportion of antibiotic ⊥prescriptions in accordance with current treatment guidelines	N: Total number of antibiotic prescriptions in accordance to UCG 2016/2023 (the Antibiotic, dose, and frequency) D: Total number of antibiotics prescriptions	Antibiotics prescribed by WHO AWaRe
Proportion of antibiotic	N: Total number of antibiotic	Antibiotics prescribed by

Indicator	Indicator definition	Disaggregation
combinations with overlapping coverage	combinations with overlapping coverage D: Total number of antibiotic prescriptions	WHO AWaRe

N – Numerator, D – Denominator

Trend analysis will be used to derive the slope coefficient for the trends in the proportion of each antibiotic and level of AMR over a five-year period for the 9 RRHs. The trends will be disaggregated by age, gender, bacterial infectious syndromes, wards, and bacteria types among others.

Objective three

Expected outcomes

The expected outcome is a machine learning model for predicting the clinical outcomes (length of stay, mortality, time to clinical improvement, mortality and disability) of patients with bacterial infectious syndromes due to resistant pathogens using clinical parameters and demographics. The model will be deployed as web or mobile device App based physician assistant that can be used as point of care aid in the management of patients with AMR.

Building the Machine Learning Algorithm

Several classification machine learning algorithms including, logistic regression (LR), artificial neural networks (ANN), support vector machines (SVM), random forest (RF), AdaBoost etc. will be trained and their performance evaluated by computing their accuracy, F1 score, precision, recall, and the area under the receiver operating curve (AUC) using a confusion matrix. The clinical outcomes of patients with bacterial resistant infectious will then be compared to those with susceptible infections using a Chi-square test for categorical variables and non-parametric or parametric tests for the continuous data depending on the distribution of the data. The pipeline and derived algorithms will be stored on secure servers with access limited to only authorized personnel using the RBAC model.

Objective four

Expected outcomes

The main outcome of this COI analysis will be the cost per AMR case from societal and payer (government) perspectives.

Analysis of outcome variables

Data to parameterize the model will be obtained from the data warehouse and published grey literature. The activity-based (micro-costing) technique will be used to estimate the cost of diagnosis using the cost information in records related to the procurement of medicines, including antibiotics, AMR-related investigations, and clinical care of patients with AMR at the nine surveillance sites.

The micro-costing technique decompounds each service (i.e., diagnosis and treatment options) into the inputs and quantity required to provide it. Then, the best price for each input will be found and multiplied by the amount needed. The sum of all inputs provides a good estimate of the cost per intervention. Given that the analysis will be conducted from both the societal and payer (government) perspectives, direct medical costs, direct non-medical costs, and indirect (productivity) costs will be included.

Estimating the direct medical costs of AMR

To estimate direct medical costs, we will use the available data bases of AMR data at IDI and Ministry of Health (MOH). These data will aide in estimating the resource use and costs of identified AMR cases. Data on health resource use – drugs, laboratory tests and other supplies – will be obtained from the literature to estimate the type and quantity of resources, which will then be multiplied by the unit costs obtained from the available price catalogues (either from Uganda's Joint Medical Stores or management sciences for health (MSH)). Regarding personnel-related costs, we will use the opportunity cost of paid-time of all health workers. Data on revised salaries will be obtained from the MOH or Ministry of Public Service to improve the precision of wage estimates.

All records related to the procurement of medicines including antibiotics, AMR related investigations, and healthcare of patients with AMR at the nine RRH surveillance sites over the five-year period will be reviewed and data abstracted. All the volumes and related costs of medicines procured, and AMR related investigations and healthcare by the surveillance site over the evaluation period will be considered. All National Medical Stores (NMS) invoices and delivery notes for each delivery cycle over the five-year period will be reviewed and data extracted on the volume of drugs delivered to the surveillance site, issued out to the different units, and the unit cost of each medicine. Also, information on clinical care including length of stay, investigations, and other medicines will

be extracted. The parameters of interest will include volume and types of medicines and cost of the medicines, types and cost of investigations, time of stay in hospital, and other costs of healthcare of patients with AMR, and the hospital medicines, investigation, other sundries for clinical care, and total budget.

Estimating the direct non-medical costs of AMR

The overhead and recurrent costs related to out-patient and hospital treatment of AMR complications will be estimated from the available data at the MOH and IDI databases and the WHO Choosing Interventions that are Cost-effective (WHO-CHOICE) database for Uganda. All capital costs will be annualized using a discount rate of 3% with an assumed lifespan of 30 (buildings), 5 (computers) and 10 (furniture) years to estimate the actual economic/ opportunity costs. Other costs, such as transportation and upkeep for patients, will be obtained from the literature if unavailable at the MOH or IDI.

Estimating the indirect costs of AMR

Productivity losses for patients and their caregivers due to AMR-related morbidity and mortality will be estimated using the friction cost approach (FCA). The FCA estimates the friction period – the time it takes an organization to replace an absent worker due to morbidity or mortality—and largely depends on the country's unemployment rate [15, 16]. We will sum lost time spent in transit to hospitals/ health facilities (for patients and caregivers), seeking care, convalescing, and being admitted to a hospital (for patients and caregivers). The time lost will be valued at Uganda's gross domestic product (GDP) per capita as a proxy for wages, given the unavailability of these data in Uganda. The friction period accounts for absenteeism (actual absence from work) and presenteeism (available but not at full productive level, thus affecting the workflow). Given Uganda's unemployment rate of 12% [17], we will benchmark the published and grey literature to estimate the friction period for Uganda and indirect costs associated with AMR [16]. All future costs will be discounted at an annual rate of 3%, as recommended by the second panel on cost-effectiveness in health and medicine [18].

The societal cost will be the summation of direct medical costs, direct non-medical costs, and indirect costs, while the payer cost will only include direct medical and direct non-medical costs. The total cost for antibiotics and the essential medicines and other AMR related healthcare costs will be derived from the product of the volume of the items/procures consumed, duration of provision of services, and the unit cost of each.

The main study outcome will be the cost per AMR case from societal and payer (government) perspectives. This will be the sum-product of the quantity of resources needed to treat one AMR case and its unitary costs. All costs will be converted to United States dollars (USD) using the bank of Uganda official exchange rates, while costs from the literature will be inflated to the reporting calendar year using the local consumer price indices. Given the ever-present uncertainty surrounding parameter estimates, we will assign different distributions to each model parameter using the 95% confidence intervals and standard errors if available and, if unavailable, using a ± 50% range. Monte Carlo simulation will be used to generate 1000 iterations of the model results; these new estimates will provide the cost of an average case of AMR with a 95% credibility range around the estimated cost. We will also perform a one-way sensitivity analysis – presented as a tornado diagram—to determine which variables greatly influence costs. All analyses will be programmed in Microsoft Excel and supported by R software (version 4.3.0).

Ethical consideration

This study has been approved by the Infectious Diseases Institute Research Ethic Committee (IDI-REC-2023-67:), which also granted a waiver of consent. The study has also been approved by the Uganda National Council for Science and Technology (UNCST - HS3690ES). No participants' identification information will be used in dissemination or publication of the study results.

Results

Once implemented, the expected results from the study will include a robust and dynamic data warehouse, AMR and AMU trends, a machine learning pipeline to model the AMR-related clinical outcomes, and the cost per AMR case to give indication of the AMR-related economic burden.

Discussion

The study will create a data warehouse and analyze AMR and AMUC rates, trends using data science and traditional statistical approaches. We will also determine the associated societal costs in Uganda and develop a machine learning model for predicting AMR related clinical outcomes. The dearth of data on AMU and AMR trends, clinical outcomes and the economic burden of AMR on the healthcare system is a major hinderance to policy formulation, stewardship interventions and resource allocation.

The proposed data warehouse will help organize the AMR and AMUC data making it accessible and linkable to other databases to allow deeper data mining to answer research and program questions through individual and collaborative research. This will contribute to one of WHO's AMR surveillance plan pillar of generating knowledge through generation of quality data and evidence [19]. On the other hand, a concise description of the AMR and AMU trends is critical to inform stewardship interventions to combat the escalating burden of AMR. Surveillance reports estimate that the AMR burden including mortality is disproportionately borne by Sub-Saharan Africa (SSA) but this has not been fully evaluated due to the scarcity of data in the region [20]. Further, optimization of AMU is important to address one of the biggest drivers of AMR in this setting – the overuse and misuse of antimicrobials [21-23]. By exploring the AMU trends and the influencing factors, the study will inform strategies to address this problem.

Since the institution of Global Antimicrobial Resistance and Use Surveillance System (GLASS) in 2015, Uganda like many other countries in SSA have developed their NAP on AMR and started generating AMR and AMU surveillance data [24]. However, the AMR data generated has been lacks data on the certain aspects of the AMR burden including the clinical outcomes (e.g. disability and mortality) and the AMR economic burden [25]. Moreover, the few reports indicate that the DALYS in those who survive AMR are increased [9], and might equal to those of influenza, tuberculosis and HIV combined [25]. This limits the optimization of clinical care for AMR patients as well national planning of AMR programs and guiding resources allocation. The study will employ machine learning algorithms and cost of illness estimation approaches to fill these critical gaps in the Uganda AMR surveillance data.

Our study's strength lies in focusing on critical gaps currently pegging back AMR surveillance in SSA and employing a multipronged approach to address the gaps including data science and economic evaluation approaches. This will provide information based on a wide and robust evidence base to inform AMR clinical management and control policies. The study is limited in the qualitative assessment of the gaps, which would have provided a comprehensive overview of the drivers of AMR and better inform the policy formulation process. Further, the use of a retrospective data will pose challenges of missing variables and inability to control for confounders for the observed outcomes while the economic evaluation will impute some variables from literature that are not routinely collected, which may bias some of the conclusions.

In conclusion, the data warehouse will promote access to AMR and AMU data to answer AMR program and research questions using a wide evidence base. The AMR-related clinical outcomes and

AMR economic burden data will facilitate improvement in clinical management of AMR patients and guide resource allocation to support AMR surveillance and interventions.

Acknowledgements

We would like to thank the study investigators in review and approve of the manuscript for publication.

Conflict of interest

The authors report no conflict of interest

Funding source

The study will be funded by a grant from the Wellcome Trust through the CAMO-NET research group (Award number 226692/Z/22/Z).

Abbreviations.

AAS: Annual Antibiotic Surveys
ACE: African Center of Excellence in Bioinformatics & Data-Intensive Sciences
AMR: Antimicrobial Resistance
AMUC: Antimicrobial Use and Consumption
AST: Antimicrobial Susceptibility Test
AUC: Area under the curve
AWaRe: Access, Watch, Reserve
CAMO-NET: The Centers For Antimicrobial Optimization Network
COI: Cost of Illness
DALYs: Disability Adjusted Life Years
FCA: Friction cost approach
GLASS: Global Antimicrobial Resistance and Use Surveillance System
HIV: Human Immunodeficiency Virus
HPC: High-Performance Scientific Computing Infrastructure
IDI: Infectious Diseases Institute
MOH: Ministry of Health
MSH: Management Sciences for Health
NAP: National Action Plan

NMS: National Medical Stores

ODK: Open data Kit

PPS: Point Prevalence Surveys

RBAC: Role-based access control

RRH: Regional Referral Hospital

SSH: Secure Shell Protocol

SSA: Sub-Saharan Africa

UCG: Uganda Clinical Guidelines

URTI: Upper Respiratory Tract Infection

USD: United States Dollars

WHO: World Health Organization

WHO-CHOICE: WHO Choosing Interventions that are cost-effective

References

1. WHO, *Antimicrobial resistance.* [Internet], Nov 2021.
2. Dadgostar P, *Dadgostar P. Antimicrobial resistance: implications and costs* Infect Drug Resist, 2019. **12**: p. 3903–3910.
3. Murray C J L, et al., *Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis.* The Lancet, 2022 Feb. **399**(10325): p. P629-655.
4. Elton L, et al., *Antimicrobial resistance preparedness in sub-Saharan African countries.* BMC, 2020. **145**(9).
5. Dever LA and Dermody TS, *Mechanisms of bacterial resistance to antibiotics.* Arch Intern Med, 1991 May. **151**(5): p. 886-95.
6. Čižman M, et al., *Correlation between Antibiotic Consumption and Resistance of Invasive Streptococcus pneumoniae.* Antibiotics (Basel), 2021 Jul. **10**(7): p. 758.
7. WHO, *World Health Organization (2012a) Factsheet No: 194: Antimicrobial resistance. March 2012.* 2012 March.
8. WHO, *Antimicrobial resistance.* [Internet], 2021 Nov.
9. Global Health CEA, *The DALYs.* [Internet].

10. Harris PA, et al., *The REDCap consortium: Building an international community of software partners.* J Biomed Inform, 2019 May 9.
11. Kiggundu R, et al., *Point Prevalence Survey of Antibiotic Use across 13 Hospitals in Uganda.* Antibiotics (Basel), 2022 Feb 4. **11**(2): p. 199.
12. Riley RD, et al., *Calculating the sample size required for developing a clinical prediction model.* BMJ, 2020. **368**: p. m441.
13. Goldenholz DM, et al., *Sample Size Analysis for Machine Learning Clinical Validation Studies.* Biomedicines, 2023. **11**: p. 685.
14. GHCC, *Reference Case for Estimating the Costs of Global Health Services and Interventions.* [Internet]: p. B-4.
15. Brouwer W and van Baal P, *Moving Forward with Taking a Societal Perspective: A Themed Issue on Productivity Costs, Consumption Costs and Informal Care Costs.* Pharmacoeconomics, 2023: p. 0123456789.
16. Hanly P, et al., *Estimating Global Friction Periods for Economic Evaluation: A Case Study of Selected OECD Member Countries.* Pharmacoeconomics, 2023: p. 0123456789.
17. UBOS, *UBOS. National laborforce survey 2021*. [Internet]. **2022**.
18. Sanders GD, et al., *Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second panel on cost-effectiveness in health and medicine.* JAMA, 2016. **316**(10): p. 1093–103.
19. World Health Organization, *Sixty-Eighth World Health Assembly* [Internet], 2015.
20. Kariuki S, et al., *Antimicrobial Resistance Rates and Surveillance in Sub-Saharan Africa: Where Are We Now?* . Infect Drug Resist, 2022 Jul **15**: p. 3589-3609.
21. Carl Llor and Lars Bjerrum, *Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem.* Ther Adv Drug Saf, 2014 Dec. **5**(6): p. 229–241.
22. Afari-Asiedu S, et al., *To sell or not to sell; the differences between regulatory*

and community demands regarding access to antibiotics in rural Ghana. Journal of Pharmaceutical Policy and Practice 2018. **11**(30).

23. Sakeena MHF, Bennett AA, and McLachlan AJ, *Enhancing pharmacists' role in developing countries to overcome the challenge of antimicrobial resistance: a narrative review.* Antimicrob Resist Infect Control, 2018 May. **7**: p. 63.
24. Nabadda S, et al., *Implementation of the World Health Organization Global Antimicrobial Resistance Surveillance System in Uganda, 2015-2020: Mixed-Methods Study Using National Surveillance Data.* JMIR Public Health Surveill. , 2021 Oct. **7**(10): p. e29954.
25. Cassini A, Högberg LD, and Plachouras D, *Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis.* Lancet Infect Dis, 2019 Jan. **19**(1): p. 56–66.