MBIRA Statistical analysis plan
Version 1.1
Dated: Feb 2022

Authors: Andrea Rehman (study statistician), Alexander Aiken (study PI),

Marlieke de Kraker (external epidemiological advisor)

Table of Contents

1. Background

The statistical analysis plan for the MBIRA (Mortality from Bacterial Infections Resistant to Antibiotics) observational cohort study is presented here. The plan defines *a priori* the analysis which will be completed for the primary and secondary outcomes of the study, including sub-group analyses. The plan adheres to STROBE guidelines (von Elm et al., 2007) for the reporting of observational studies. This statistical analysis plan (SAP) has been finalised prior to the completion of data collection and before statistical analysis has been undertaken.

The burden of antimicrobial resistance (AMR) and linkage of resistance status to clinical outcomes in low-and middle-income countries forms a research gap in our understanding of the impact of AMR. A large study in European adults found a significant burden of morbidity and mortality was attributable to third-generation-cephalosporin (3GC) resistant *E. coli* blood stream infections (de Kraker et al., 2011). The generalisability of these findings to LMIC settings is largely unknown. Therefore, among inpatients with bacteraemia caused by species within the Enterobacterales genus identified by blood culture, we aimed to quantify the relationship between third generation cephalosporin resistance status and clinical outcomes including mortality and length of hospital stay in 8 African hospitals.

The study is funded by the Bill and Melinda Gates Foundation, grant reference number INV-003519.

2. Design

This is an observational matched parallel cohort study carried out in 8 African hospitals. It included inpatients with Enterobacterales-species bacteraemia and their uninfected matches, matched by calendar date (ideally +/- 2 weeks of admission date, longer periods accepted if needed), ward and of comparable age and whose inpatient stay was at least as long as the time between hospital admission and positive blood culture in the bacteraemia case. Bacteraemia patients and matches were recruited in a 1:2 ratio and were followed to determine clinical hospital outcomes 30 days after a positive blood culture (or for matches, enrolment to the study) and at discharge from hospital. The study aimed to recruit 120 bacteraemia inpatients per site and commenced recruitment in November 2019. Recruitment will stop in January 2022 as planned.

Objectives

a) Primary objective: To quantify the association between AMR status (defined in terms of 3GC resistance status) and hospital mortality rate among inpatients with proven Enterobacterales-species bacteraemia in African hospitals
b) Secondary objective: To quantify the association between AMR status and other clinical outcomes, including 30-day mortality and length of hospital stay
c) Secondary objective: To explore (through sub-group analysis) if the association between AMR status and clinical outcomes varies by i) site ii) age group iii) species of bacteraemia (based on species with sufficient numbers, likely to be only *Klebsiella pneumoniae* and *E. coli*) and iv) community-acquired v hospital-acquired infections.
d) Secondary objective: as an alternative form of AMR exposure status (in place of 3GC-resistance status), we will use the concordancy of antibiotic treatment (ie concordance of drugs used against the causative pathogen based on antimicrobial susceptibility testing results) in the first 48 hour period of treatment of the bacteraemia to determine impact of AMR on hospital mortality, 30-day mortality, and length of hospital stay.
e) Secondary objective: as additional analyses within the “concordancy of antibiotic treatment” detailed in part (d), we will examine specific subgroups of pharmacological-microbiological

interest (AmpC-producers, ESBL imputation decisions[1], BLBI v carbapenem impact, inadequate dosing), if suitable numbers are available.

3. Recruitment

The flow of participants through the study will be illustrated using a diagram as per STROBE guidelines. Exclusions with reasons, losses to follow-up and numbers analysed will be displayed.

Participating sites were instructed to enrol all consecutive bacteraemia patients over a continuous recruitment in 2020-22. Matching patients were selected at random from suitable hospital inpatients. Investigators were advised to do this using a random number generator software downloaded onto a portable device – we suggested use of the "Random: All Things Generator" App. Participating patients gave informed consent, according to local protocols.

4. Characteristics of the study population

Characteristics of participants, obtained at enrolment will be described by AMR status and infection status (bacteraemia v match) according to the analysis principles already described.

5. Study definitions

3rd generation cephalosporin resistance status will be determined as resistant to any of ceftriaxone, cefotaxime or cefpodoxime. If there are conflicting results between the reference laboratory (Stellenbosch University, South Africa) and the original study site, the reference laboratory results will be applied. Susceptibility testing results with an "intermediate" or "therapeutic uncertainty" will be grouped with "resistant" isolates.

For generation of the "antibiotic concordancy" status of BSI treatment, we will follow approaches described in more detail in the relevant study training guide (document titled "MBIRA study: Guide to scoring appropriate-ness of antibiotic use by pathogen, including imputation of missing antibiotic susceptibilities", version 14th December 2020). Briefly, this is a parameter that considers all the different antibiotics used in the first 48 hours of treatment (approximated by considering day 0, day 1 and day 2, where day 0 represents the day the blood culture was taken) against all the relevant laboratory susceptibility testing results. The parameter is graded into the following categories for each drug: "Yes = concordant", "yes, but inadequate dosing", "no = non-concordant" and "unable to determine". Documented receipt of one (or more) doses of one (or more) antibiotic(s) with a "yes" categorization in this day 0 to day 2 period would be considered "concordant" initial treatment. There are liable to be a relatively small number of individuals where there are only antibiotic treatments that are "unable to determine" or "inadequate dosing" status for concordancy – these BSI patients (and their corresponding matches) will be excluded from analysis.

6. Sample size considerations

This study is powered to test the hypothesis that there is a difference in the Odds Ratio for death between **3GC-R** bacteraemia cases and their respective controls (=OR_1) as compared to **3GC-S** bacteraemia cases and their respective controls (=OR_2). If the confidence interval of the ratio of these two Odds Ratios (ie OR_2/ OR_1) includes 1.0, we cannot exclude the possibility that there is no true difference in the ratios. We therefore aimed to have a sufficiently large study such that

[1] This means, inference of non-effective-ness of antibiotics in other antibiotic classes (eg. β-lactam-β-lactam-inhibitor (BLBI) drugs such as co-amoxiclav) based on laboratory evidence of the presence of an ESBL enzyme.

plausible estimates for both OR_1 and OR_2 will combine to give a ratio of ratios with a 95% confidence interval that excludes the value of 1.0.

We estimated the expected 95% confidence interval of the ratio of ratios (ie OR_2/ OR_1) according to the following formula:

For OR_1 with standard error $SE(OR_1)$ and OR_2 with standard error $SE(OR_2)$, then the ratio OR_2/OR_1 has SE(ratio) = $\sqrt{SE(OR_1)^2 + SE(OR_2)^2}$ (Altman and Bland, 2003).

Projected values

Calculations are performed for all bacterial species grouped together as *Enterobacterales*. Sub-group analyses restricting to *E. coli* and *K. pneumoniae* (the two most common individual species) or to particular age groups are also planned, but the study was not powered for those sub-analyses. Based on a recruitment period of two years, we expected a total of 1,200 bacteraemia cases across 10 sites with 2 non-infected matches / bacteraemia patient. Assuming a 3GC resistance rate of 20% this would result in the following cohorts:

3GC-S bacteraemia cohort – 960 patients, 1720 matches; OR_1 = 2.0, 95%CI 1.6-2.5

3GC-R cohort – 240 patients, 480 matches; OR_2 = 6.1, 95%CI 4.1-9.0

Using the above formula with these expected values, gives a ratio of odds ratios of 3.0 with a projected 95%CI of 1.90 to 4.76 for 1,200 bacteraemia cases. This projected confidence interval excludes the value of 1.0 with a substantial margin of error. We therefore considered this to be adequate powering of the study design.

7. Outcome variables

For all outcomes, the timescale begins at the date of blood culture or, for matches, the equivalent day of inpatient stay as their matched bacteraemia case.

The primary outcome is the rate of in-hospital mortality compared between the cohort comprising 3GC-resistant BSI and their matches versus the cohort comprising 3GC-susceptible BSI and their matches.

Secondary outcomes are:

a) 30-day mortality including follow-up beyond hospital discharge
b) Length of inpatient stay (after blood culture or enrolment)[2]

8. Analysis principles

Analysis principles closely follow the published GLASS methodology (WHO team: Global Antimicrobial Resistance Surveillance System (GLASS), 2020). Primary and secondary outcomes from enrolled participants who meet eligibility criteria will be analysed with the aim of obtaining estimates and 95% confidence intervals of the effect of AMR status on outcomes. Matching will be accounted for using robust standard errors and we will consider clustering at the site-level in all regression analyses.

[2] A small number of patients are expected to have their hospital outcome recorded as "transferred to another hospital" or "left hospital against medical advice" (ie abscond from hospital). For these patients, we will exclude them from the LOS analysis (ie consider as lost to follow-up) as their "ideal" dates of leaving healthcare facility are unknown.

We will use data from all enrolled patients in the analysis, except for matching patients who are found to not meet matching criteria, and subsequent bacteraemia cases with no remaining matches.

An assessment of whether contributing patients differ from non-contributing patients will be made to explore the potential for selection bias in the sample. We will not conduct statistical tests comparing bacteraemia patients to their matches.

All analyses will be adjusted for categorical age group (neonate, infant, child, adult). To assess for the effect of confounding, whereby the matched uninfected group may differ in their risk (of death) independent of an AMR BSI, we will adjust for potential confounding factors. Factors will include the Charlson co-morbidity index, HIV status, the presence of indwelling devices and other factors based on clinical knowledge and/or associated with both outcome and AMR status. Collinearity will be assessed by variance inflation factor or correlation.

We will not adjust our analyses for bacterial species as we consider this to be a characteristic of the infection and hence there is no corresponding data for the matching patients. We will however perform sub-group analysis by bacterial species, as described in objective c)iii) above.

9. Missing data

The amount (number, percentage) of data that is missing will be reported for each of the main outcome and potential confounding variables. Characteristics of those with and without missing values will be compared to explore whether a missing at random assumption might be valid. T-tests or linear regression will be used for normally distributed continuous covariates comparing the mean values of those with and without missing data, Mann-Whitney nonparametric test for continuous skewed covariates, and chi squared tests and percentages for categorical covariates.

Possible reasons for missing data in general and any implications for interpretation of study findings will be discussed. An analysis assuming missing at random (Sterne et al., 2009) may be carried out, where covariates associated with missing data (as identified above) will be included in adjusted models and/or multiple imputation will be used. Multiple imputation with chained equations will generate 10 imputation datasets. Imputation models will include outcomes, site, resistance status, species, age category, Charlson Co-morbidity Index category, HIV status and any other factors associated with substantial amounts of missing data. Imputation datasets will be combined using Rubin's rule to generate overall estimates.

10. Analysis methods

For continuous variables with normally distributed residuals, the mean and standard deviation will be presented. For skewed continuous variables, either geometric mean or the median and inter-quartile range (IQR) will be presented, and for categorical variables the number, total, and percentage in each category. For event data, the number of events, person-time at risk and rate will be presented. Significance tests will be two-sided with 5% level of significance and reported using overall Wald p-values in statistical models which incorporate robust standard errors.

Effects attributable to third-generation cephalosporin resistance will be determined as the ratio of effect measures for the resistant and susceptible cohorts, with their estimates and 95% CI calculated (Altman and Bland, 2003). Both minimally adjusted and fully adjusted models will be presented for the primary analysis. All analysis will be performed for bacteraemia cases with their matched patients separately for 3GC resistant and susceptible cohorts.

For analysis of hospital mortality, cumulative incidence plots based on Aalen-Johansen estimators will be used for graphical presentation, incorporating the competing risk of discharge from hospital. Cause-specific Cox regression will be used to estimate the direct impact of bacteraemia on hospital mortality, censoring hospital discharge, and to estimate the indirect impact of bacteraemia on hospital mortality by modelling discharge as an outcome, while censoring hospital mortality. Secondary analysis will estimate sub-distribution hazard ratios considering mortality as the outcome of interest and considering discharge as a competing event (Fine and Gray, 1999). Results will be presented as hazard ratios with 95% confidence intervals. Follow-up time will stop at death, loss to follow-up or hospital discharge and will be censored at 200 days post-enrolment. Any participant with an event (or who is censored) on the first day of follow-up will be included in analysis by setting their time to half a day. The time scale used for time to event analysis will be time since blood culture (or comparable day of inpatient stay for matches).

The 30-day mortality outcome will be analysed using Generalised Linear Models with Poisson distribution and log link to generate relative risk ratios and their 95% CI (Zou, 2004). Uncontactable patients will be excluded.

Excess length of stay will be analysed using Generalised Linear Models with Gamma distribution and log link to generate relative ratios and their 95% CI excluding the small number of patients who left hospital against medical advice or were transferred out.

Pre-specified subgroup analysis of hospital mortality, 30-day mortality and excess length of stay will be carried out by (i) site and (ii) age category (iii) bacterial species and (iv) community-acquired v hospital-acquired infection. All sub-group analyses will be presented with a fully adjusted analysis only. Effect modification will be examined by incorporating an interaction term between subgroups and case or match status separately within the resistant and susceptible cohorts. Forest plots will be used to visually display stratified estimates of effect. In the case of low numbers of events which make stratum specific estimates inestimable, those strata will either be combined with others or excluded from forest plots and analysis.

In an exploratory analysis, we will perform a test of association between 3GC-resistance status and concordancy of initial antibiotic treatment with relative risk estimates to measure association. We will adjust this analysis for categorical age group only, as we anticipate data sparsity will preclude further adjustment.

As described above, we will also perform analyses replacing 3GC status with “concordancy of antibiotic treatment in first 48hr period”. This will be a binary variable created based on known antibiotic treatments received in this time period. We will exclude BSI patients who either received no antibiotic treatments at all, no antibiotics treatments of a known effective dose or for whom we are unable to determine the concordancy of an antibiotic treatment. For this analysis, we will also perform a sensitivity analysis in which we exclude patients known to have died during this first 48 hr period as we believe that these “early” deaths may reflect patients who have not had sufficient time to benefit from treatment. We will also perform additional sensitivity analyses for various pharmacological-microbiological sub-groups (see Appendix 2 of antibiotic appropriate-ness guide), if sufficient data available. These are intended to determine if alternative microbiological assumptions would change the major findings of the analysis.

References

ALTMAN, D. G. & BLAND, J. M. 2003. Interaction revisited: the difference between two estimates. *BMJ,* 326**,** 219.

DE KRAKER, M. E., WOLKEWITZ, M., DAVEY, P. G., KOLLER, W., BERGER, J., NAGLER, J., ICKET, C., KALENIC, S., HORVATIC, J., SEIFERT, H., KAASCH, A., PANIARA, O., ARGYROPOULOU, A., BOMPOLA, M., SMYTH, E., SKALLY, M., RAGLIO, A., DUMPIS, U., MELBARDE KELMERE, A., BORG, M., XUEREB, D., GHITA, M. C., NOBLE, M., KOLMAN, J., GRABLJEVEC, S., TURNER, D., LANSBURY, L. & GRUNDMANN, H. 2011. Burden of antimicrobial resistance in European hospitals: excess mortality and length of hospital stay associated with bloodstream infections due to Escherichia coli resistant to third-generation cephalosporins. *J Antimicrob Chemother,* 66**,** 398-407.

FINE, J. P. & GRAY, R. J. 1999. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc,* 94**,** 496-509.

STERNE, J. A., WHITE, I. R., CARLIN, J. B., SPRATT, M., ROYSTON, P., KENWARD, M. G., WOOD, A. M. & CARPENTER, J. R. 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ,* 338**,** b2393.

VON ELM, E., ALTMAN, D. G., EGGER, M., POCOCK, S. J., GOTZSCHE, P. C., VANDENBROUCKE, J. P. & INITIATIVE, S. 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet,* 370**,** 1453-7.

WHO TEAM: GLOBAL ANTIMICROBIAL RESISTANCE SURVELLIANCE SYSTEM (GLASS) 2020. GLASS method for estimating attributable mortality of antimicrobial resistant bloodstream infections. *In:* WORLD HEALTH ORGANISATION (ed.). Geneva, Switzerland.

ZOU, G. 2004. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol,* 159**,** 702-6.