Nasopharyngeal Resistome Study Statistical Analysis Plan V2.0

Author: Lucy O'Connor

Protocol Summary: Study Design and Research Questions

Study type: A case-control study.

Research Question: Does hospital admission and antibiotic treatment for pneumonia generate a persistent hospital-acquired nasopharyngeal resistome (NPR) in children in Blantyre, and is this associated with adverse health outcomes in a resource limited setting?

Primary Study Objectives:

1. To measure how the diversity and relative abundance of the NPR of children hospitalised with pneumonia differs from the NPR of healthy children in the community.
2. To measure how the diversity and relative abundance of the NPR of children hospitalised with pneumonia differs from the NPR of children with pneumonia treated with antibiotics in the community.
3. To investigate whether participants re-hospitalised with pneumonia within a 3-month period have a persistent hospital-acquired NPR pattern.

Secondary Study Objective:

4. To measure the impact of exposure to a hospital environment on the development of a 'hospital-acquired' NPR.

Exploratory Study Objective:

5. To measure how the diversity and relative abundance of the NPR of children presenting with pneumonia differs between those with and without prior antibiotic use, based on urine bioassay.

Methodology: We will recruit 350 children aged 12-36 months in four groups: (i) healthy children in the community, (ii) children with pneumonia in the community, (iii) children hospitalised with pneumonia and (iv) children re-hospitalised with pneumonia. This is illustrated in Figure 1 (below). This observational study will be conducted in Ndirande community, at Ndirande community healthcare centre and at Queen Elizabeth Central Hospital (QECH). Nasopharyngeal swabs (NPS) and urine samples will be taken at recruitment and follow-up study visits; NPS will undergo metagenomic sequencing to identify the nasopharyngeal microbiome (NPM) and NPR, urine samples will undergo a bioassay to detect antibiotic activity, indicating recent antibiotic exposure. Demographics, co-morbidities, pneumonia severity, antibiotic exposure and health outcomes will be recorded at study visits.

Participants will be frequency-matched for sex across the healthy community, community pneumonia, hospitalised pneumonia and re-hospitalised pneumonia groups for the statistical analysis. Figure 1 illustrates the key groups for comparison in the statistical analysis.

Children hospitalised at QECH with pneumonia (without any hospital admission in the preceding 6 months), who are not re-hospitalised within 3 months of recruitment will form the controls (see Figure 1); any hospitalised participant that is re-hospitalised will be excluded from the control group for the case-control analysis. Children admitted to QECH with pneumonia within 3 months of a previous hospital admission for pneumonia,

meningitis or an infection managed with intravenous beta-lactam and/or aminoglycoside antibiotic will form the cases.

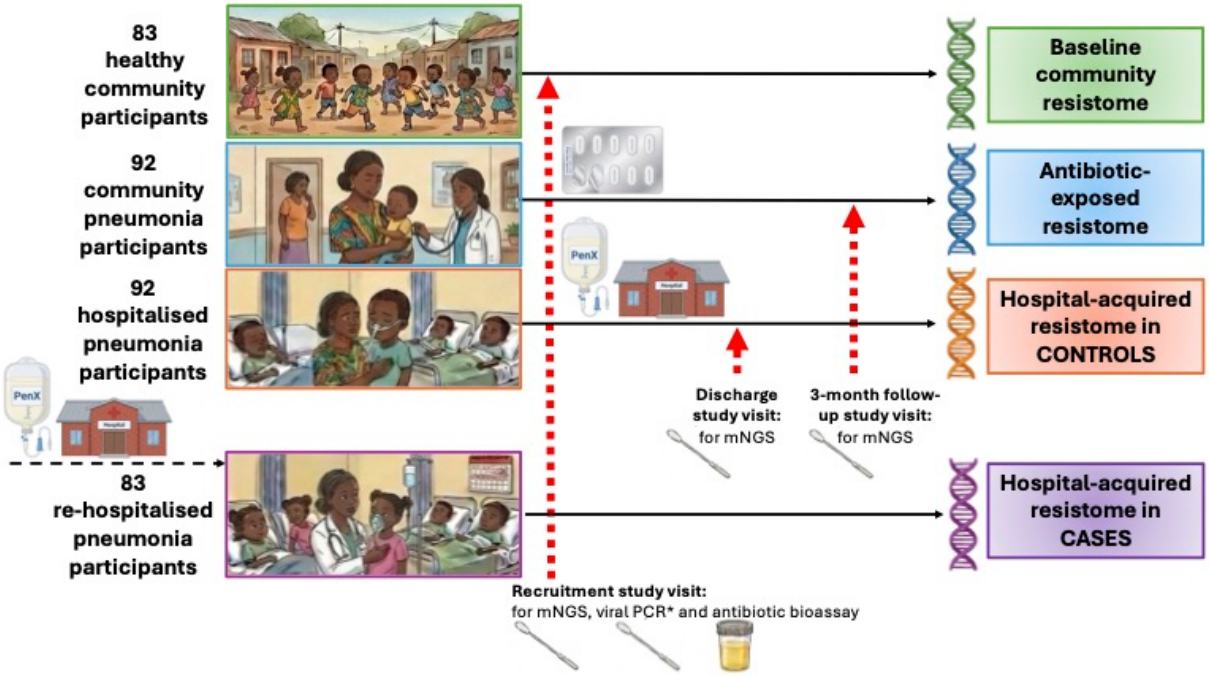

Figure 1. Study design and sampling frame. Red arrows show study visits. Small icons show antibiotic and hospital exposure. Hospitalised pneumonia participants (orange boxes) that are not re-hospitalised within 3 months of discharge form the control group for the case-control comparison. *Only participants with pneumonia will have a second NPS taken for viral PCR at recruitment. mNGS, shotgun metagenomic sequencing.

Expected results: The study is expected to describe NPR evolution following antibiotic selective pressure and residential healthcare exposure for pneumonia in children in Blantyre. It will provide contemporary data on AMR carriage in the community, hospital a re-admission rates with pneumonia and antibiotic use patterns.

Study Outcomes

Diversity is a measure used in microbial ecology to quantify the variety of microorganisms in a community. For the analysis of resistome composition, diversity is a measure of antibiotic resistance gene (ARG) richness, the number of different ARGs within a microbial community, and evenness or relative abundance of ARG distribution within the community. Alpha diversity quantifies the ARG richness and evenness within one sample or individual; beta diversity compares ARG richness and evenness between communities. In this study, alpha diversity of the NPR within each sample will be measured using the Shannon Diversity Index, a continuous metric, where a higher number indicates a greater level of ARG diversity and evenness within the sample.

The Bray-Curtis Dissimilarity Index will be calculated to quantify the beta diversity, or compositional (dis)similarity, between the NPR of the four different groups. This compares the abundance of ARGs between two samples, by considering both size (overall abundance per sample), shape (abundance of each type of ARG) and the presence or absence of ARGs within a microbial community. Values range between 0 and 1, where 0 indicates that two

samples have the same composition (they share all their ARGs) and 1 indicates the two samples do not share any ARGs.

The study will focus on two key outcomes in analysis of NPM and NPR composition post-antibiotic and hospital exposure:

- Difference in beta diversity
- Difference in relative abundance of ARG classes for first line antibiotics used to treat pneumonia

Secondary outcomes in the analysis of NPM and NPR composition post-antibiotic and hospital exposure will include:

- Difference in alpha diversity
- Difference in relative abundance of individual ARGs.

Study Time End-points

The study end-point is recruitment for community healthy and hospital readmission participants for all analyses. The study end-point for the primary analysis is at 3-months post-recruitment for community pneumonia and hospital pneumonia participants, which is likely to represent a new post-antibiotic baseline. This time-point was selected as studies of paediatric infective-asthma exacerbations have demonstrated microbiome recovery to a new baseline within 3 months (Hou et al., 2022); similarly, pneumococcal carriage surveillance revealed a multi-drug-resistant lineage in global pneumococcal sequencing cluster 1 (GPSC1) at 100 days post-antibiotic exposure, suggesting an enduring impact of antibiotic selection pressure on the resistome (Tonkin-Hill et al., 2022). The hospital readmission group represent a sampling time-point within 3-months of hospital discharge. For primary analyses, this group will be compared with the 3-month end-point in the hospital pneumonia group.

Secondary analyses will include comparison of resistome composition between the groups at the following timepoints:

- Recruitment of community pneumonia participants. The recruitment NPS in this group is likely to represent an acute 'antibiotic-exposed' resistome; 70% of patients report self-administration of antibiotics prior to seeking healthcare assessment in Malawi (Limwado et al., 2024).
- Recruitment and discharge for hospital pneumonia participants.

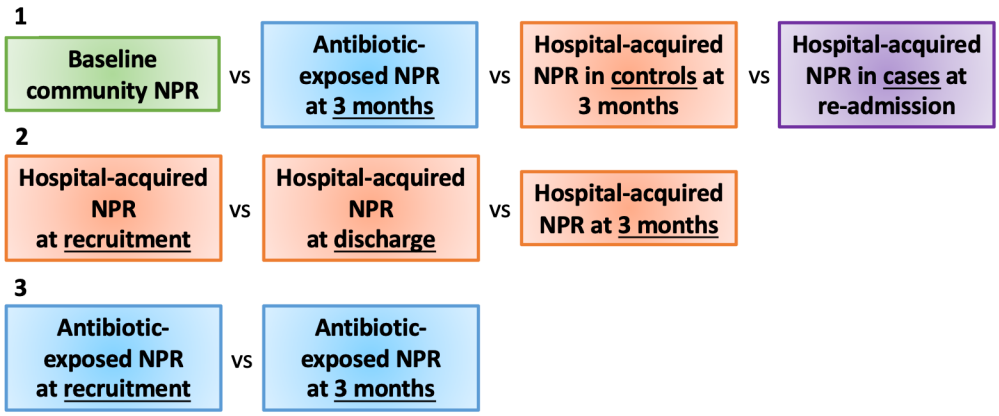

Figure 2. An illustration of the planned statistical analysis to address the primary study objectives, through comparison of NPR composition between groups, and between timepoints within groups.

Primary Outcomes

1. The composition of the nasopharyngeal resistome (NPR), as measured by the diversity and relative abundance of antibiotic resistance gene (ARG) classes detected by metagenomic sequencing of nasopharyngeal samples collected from the children hospitalised with pneumonia at recruitment, discharge and 3-month follow up.
2. The composition of the NPR, as measured by the diversity and relative abundance of ARG classes detected by metagenomic sequencing of nasopharyngeal samples collected from children treated for pneumonia in the community at recruitment and 3-month follow up.
3. The composition of the NPR, as measured by the diversity and relative abundance of ARG classes detected by metagenomic sequencing of nasopharyngeal samples collected from healthy children at recruitment.
4. The composition of the NPR, as measured by the diversity and relative abundance of ARG classes detected by metagenomic sequencing of nasopharyngeal samples collected from the children re-hospitalised with pneumonia within a 3-month timeframe, at recruitment.

Secondary Outcomes

1. The composition of the NPR, as measured by the relative abundance of individual ARGs detected by metagenomic sequencing of nasopharyngeal samples collected from the children hospitalised with pneumonia at recruitment, discharge and 3-month follow up.
2. The composition of the NPR, as measured by the relative abundance of individual ARGs detected by metagenomic sequencing of nasopharyngeal samples collected from children treated for pneumonia in the community at recruitment and 3-month follow up.
3. The composition of the NPR, as measured by the relative abundance of individual ARGs detected by metagenomic sequencing of nasopharyngeal samples collected from healthy children at recruitment.
4. The composition of the NPR, as measured by the relative abundance of individual ARGs detected by metagenomic sequencing of nasopharyngeal samples collected from the children re-hospitalised with pneumonia within a 3-month timeframe, at recruitment.
5. The composition of the nasopharyngeal microbiome (NPM), as measured by the relative abundance of bacterial species detected by metagenomic sequencing of nasopharyngeal samples collected from the children re-hospitalised with pneumonia within a 3-month timeframe, at recruitment.
6. The composition of the NPM, as measured by the relative abundance of bacterial species detected by metagenomic sequencing of nasopharyngeal samples collected from the children hospitalised with pneumonia at recruitment, discharge and 3-month follow up.
7. The composition of the NPM, as measured by the relative abundance of bacterial species detected by metagenomic sequencing of nasopharyngeal samples collected from children treated for pneumonia in the community at recruitment and 3-month follow up.

8. The composition of the NPM, as measured by the relative abundance of bacterial species detected by metagenomic sequencing of nasopharyngeal samples collected from healthy children at recruitment.
9. The presence of Respiratory Syncytial Virus, Rhinovirus, Influenza A and B in nasopharyngeal swabs from participants with pneumonia at recruitment, as detected by viral RNA polymerase chain reaction (PCR).
10. The use of antibiotics prior to recruitment, measured by urinary antibiotic activity bioassay, or (if not available) reported antibiotic use, for children treated for pneumonia in the community, children hospitalised with pneumonia and children re-hospitalised with pneumonia.
11. Severe pneumonia in hospitalised pneumonia and re-hospitalised pneumonia participants at hospital admission, as determined by documentation of the presence of cough, difficulty breathing and any of the following:
    - Central cyanosis or oxygen saturations < 90%
    - Inability to feed, convulsions, lethargy or impaired conscious level.
12. Health outcomes, measured by documented or reported hospitalisation, healthcare clinic attendance, antibiotic prescriptions and mortality at 3-month follow up of children treated for pneumonia in the community and children hospitalised with pneumonia.
13. Reported duration of hospital admission for children hospitalised with pneumonia, measured in days, at the discharge study visit.
14. Reported duration of shared bed-spaces during hospitalisation for children hospitalised with pneumonia, measured in days, at the discharge study visit.

Quality Control for Metagenomic Sequencing Data

There is no published gold standard for quality control of shotgun metagenomic sequence data. The following pre-defined analysis thresholds will be applied to bacterial reads in the metagenomic sequence data to ensure high-confidence alignments to taxa present in the NCBI database:

1) >90% identity matches of reads and assembled contigs to reference sequences.
2) Alignment lengths $\geq$ 50bp to reduce false positive alignments.

Reads will be normalised using Reads Per Kilobase per Million (RPKM) mapped reads, permitting quantitative comparison of the relative abundance of bacterial and antibiotic resistance gene (ARG) reads between samples, by dividing both transcript length and total number of reads. RPKM is calculated by dividing the number of reads mapping to a particular gene by its length in kilobases, then multiplying by 1,000,000 and dividing by the total number of mapped reads in the sample; this adjusts for biases that may arise during sequencing and analysis, including differences in gene length and sequencing depth (Zhang et al., 2022). Normalised reads will undergo a centred log-ratio transformation to account for compositionality prior to differential abundance comparisons across groups (Aitchison, 1982).

A minimum of three types of negative control will be included in the study design to permit monitoring of contaminants throughout sample collection and processing: (i) sampling blank controls, to detect contaminant DNA introduced during the sampling procedure, (ii) DNA extraction blank controls, to monitor contaminant DNA content in extraction kits and (iii) no-template amplification controls, to monitor contaminant DNA

present in reagents and laboratory environment during library preparation and sequencing (Eisenhofer et al., 2019).

Primary Study Endpoints:

1. The difference in beta diversity of ARG classes in the NPR, as measured by the Bray Curtis dissimilarity index, and difference in relative abundance of ARG classes in the NPR between healthy community and hospital readmission groups at enrolment, and community pneumonia and hospital pneumonia groups at the 3-month study end-point.
2. The difference in beta diversity of ARG classes in the NPR, as measured by the Bray Curtis dissimilarity index, and difference in relative abundance of ARG classes in the NPR between recruitment, discharge and 3-month study time-points in the hospital pneumonia group.
3. The difference in beta diversity of ARG classes in the NPR, as measured by the Bray Curtis dissimilarity index, and difference in relative abundance of ARG classes in the NPR between recruitment and the 3-month study end-point in the community pneumonia group.

Secondary Study Endpoints: NPR Analysis

1. The difference in alpha diversity, measured by the Shannon diversity index, and relative abundance of individual ARGs in the NPR between healthy community and hospital readmission groups at enrolment, and community pneumonia and hospital pneumonia groups at the 3-month study end-point.
2. The difference in alpha diversity, measured by the Shannon diversity index, and relative abundance of individual ARGs in the NPR between recruitment, discharge and 3-month study time-points in the hospital pneumonia group.
3. The difference in alpha diversity, measured by the Shannon diversity index, and relative abundance of individual ARGs in the NPR between recruitment and the 3-month study end-point in the community pneumonia group.

Secondary Study Endpoints: NPR Analysis (Dependent on primary endpoint analysis)

4. *(Dependent on primary study endpoint 2) If there is a change in NPR composition between discharge and the 3-month study end-point in the hospital pneumonia group:* The difference in the following measures of NPR composition between the hospital pneumonia group at discharge, the healthy community and hospital readmission groups at enrolment, and the community pneumonia group at the 3-month study end-point:
   a) ARG richness and evenness (alpha diversity) of the NPR
   b) Overall community composition (beta diversity) of the NPR
   c) Relative abundance of ARG classes in the NPR
   d) Relative abundance of individual ARGs in the NPR
5. *(Dependent on primary study endpoint 3) If there is a change in NPR composition between enrolment and the 3-month study end-point in the community pneumonia group:* The difference in the following measures of NPR composition between the community pneumonia group at enrolment, the healthy community and hospital readmission groups at enrolment, and the hospital pneumonia group at the 3-month study end-point:

a) ARG richness and evenness (alpha diversity) of the NPR
b) Overall community composition (beta diversity) of the NPR
c) Relative abundance of ARG classes in the NPR
d) Relative abundance of individual ARGs in the NPR

6. *(Dependent on primary study endpoints 2 and 3) If there is a change in NPR composition between discharge and the 3-month study end-point in the hospital pneumonia group, and a change in the NPR composition between enrolment and the 3-month study end-point in the community pneumonia group:* The difference in the following measures of NPR composition between the healthy community, community pneumonia and hospital readmission groups at enrolment, and the hospital pneumonia group at discharge:
   a) ARG richness and evenness (alpha diversity) of the NPR
   b) Overall community composition (beta diversity) of the NPR
   c) Relative abundance of ARG classes in the NPR
   d) Relative abundance of individual ARGs in the NPR

Secondary Study Endpoints: NPM Analysis

7. The difference in the following measures of NPM composition between the healthy community and hospital readmission groups at enrolment, and community pneumonia and hospital pneumonia groups at the 3-month study end-point:
   a) Species richness and evenness (alpha diversity) of the NPM,
   b) Overall taxonomic composition (beta diversity) of the NPM
   c) Relative abundance of bacterial taxa in the NPM
8. The difference in the following measures of NPM composition between recruitment, discharge and 3-month study time-points in the hospital pneumonia group:
   a) Species richness and evenness (alpha diversity) of the NPM
   b) Overall taxonomic composition (beta diversity) of the NPM
   c) Relative abundance of bacterial taxa in the NPM
9. The difference in the following measures of NPM composition between recruitment and the 3-month study end-point in the community pneumonia group:
   a) Species richness and evenness (alpha diversity) of the NPM
   b) Overall taxonomic composition (beta diversity) of the NPM
   c) Relative abundance of bacterial taxa in the NPM

Secondary Study Endpoints: Health Outcomes

10. The difference in the incidence of adverse health outcomes between the community pneumonia group and hospital pneumonia group at the 3-month study end-point:
    a) Repeat healthcare clinic attendances
    b) Repeat antibiotic prescriptions
    c) Hospitalisation
    d) Mortality

Exploratory Study Analyses

11. To estimate the difference in beta diversity of the NPR between healthy community and hospital readmission groups at enrolment, and community pneumonia and hospital pneumonia groups at the 3-month study end-point, as measured by the Bray Curtis dissimilarity index determined from ARG classes, using hierarchical clustering.

12. To estimate the association between NPR composition, as defined by hierarchical clustering of Bray-Curtis dissimilarities based on ARG classes, and recurrent hospitalisation with pneumonia.
13. To estimate the association between NPR composition at discharge and recurrent hospitalisation within 3 months with pneumonia among hospitalised participants.
    a) When NPR composition is determined by ARG richness and evenness (alpha diversity) of the NPR
    b) When NPR composition is defined by hierarchical clustering of Bray-Curtis dissimilarities based on individual ARGs
    c) When NPR composition is defined by the relative abundance of ARGs to beta-lactam antibiotics
14. The difference in the following measures of NPM composition at recruitment between all hospitalised participants with pneumonia when compared to severe pneumonia:
    a) Species richness and evenness (alpha diversity) of the NPM
    b) Overall taxonomic composition (beta diversity) of the NPM
    c) Relative abundance of bacterial taxa in the NPM
15. The difference in the following measures of NPR composition at recruitment between all hospitalised participants with pneumonia when compared to severe pneumonia:
    a) ARG richness and evenness (alpha diversity) of the NPR
    b) Overall community composition (beta diversity) of the NPR
    c) Relative abundance of ARG classes in the NPR
    d) Relative abundance of individual ARGs in the NPR
16. The difference in the following measures of NPM composition between all hospitalised pneumonia participants with and without viral pathogens detected at recruitment:
    a) Species richness and evenness (alpha diversity) of the NPM
    b) Overall taxonomic composition (beta diversity) of the NPM
    c) Relative abundance of bacterial taxa in the NPM
17. The difference in the following measures of NPR composition between all hospitalised pneumonia participants with and without viral pathogens detected at recruitment:
    a) ARG richness and evenness (alpha diversity) of the NPR
    b) Overall community composition (beta diversity) of the NPR
    c) Relative abundance of ARG classes in the NPR
    d) Relative abundance of individual ARGs in the NPR
18. To estimate how the duration of hospitalisation will impact the following measures of NPR composition at discharge:
    a) ARG richness and evenness (alpha diversity) of the NPR
    b) Overall community composition (beta diversity) of the NPR
    c) Relative abundance of ARG classes in the NPR
19. To estimate how the use of shared bed-spaces during hospitalisation will impact the following measures of NPR composition at discharge:
    a) ARG richness and evenness (alpha diversity) of the NPR
    b) Overall community composition (beta diversity) of the NPR
    c) Relative abundance of ARG classes in the NPR

20. To estimate how prior antibiotic use, detected by urine bioassay, will impact the following measures of NPR composition at recruitment within each study group:
    a) ARG richness and evenness (alpha diversity) of the NPR
    b) Overall community composition (beta diversity) of the NPR
    c) Relative abundance of ARG classes in the NPR
21. To estimate the change in the relative abundance of mobile genetic elements (MGE) in the NPR of the hospital pneumonia group at enrolment and discharge study visits.

Sample Size

In total we aim to recruit 350 study participants: 83 healthy community participants, 92 participants with pneumonia in the community, 92 participants admitted to hospital with pneumonia and 83 participants re-admitted to hospital with pneumonia.
There is no established method for sample size calculation in resistome studies; we use an approach for microbiome studies (Ferdous et al., 2022). Accepting 90% power, 5% type I error (95% confidence interval) and the assumptions below, we've calculated that we will require a sample size of 83 participants in each group to detect a medium effect size (Cohen's D = 0.505), using a standard t test sample size calculation formula $n=2(Z_{1-\alpha/2}+ Z_{1-\beta})^2/\Delta^2$ as recommended for microbiome studies (Ferdous et al., 2022). Study groups with a sample size of 92 participants permit a 10% loss to follow up.

Assumptions:

- Differences in gut plasmidome composition have been detected with 41 participants per group (Willmann et al., 2019), but NPR has a lower biomass than gut resistome, therefore the resistome is likely to be less diverse, with a lower relative abundance of ARGs.
- We have used the measured difference in relative abundance of a specific macrolide ARG (ermX) 3.5 years after mass-azithromycin administration (see table 1, unpublished data) as a medium effect size that maybe clinically relevant (Δ= 0.505).
- No adjustment for level of significance for multiple comparisons has been made. However, a false discovery rate will be set using the Benjamini-Hochberg method.

Table 1. Data on which the sample size calculation is based, showing increase in the relative abundance of ermX, a macrolide antibiotic resistance gene, in children 3.5 years post azithromycin mass-drug administration. Reads per million, RPM (measure of relative abundance of an ARG, normalised for the number of DNA reads in the sample)

ermX	Mean RPM*	Standard deviation	Effect size
Azithomycin exposed group	0.34	0.10	0.505
Placebo exposed group	0.29	0.07	

Statistical approach

General Considerations

The reporting of this study will be prepared in accordance with the STROBE guidelines (von Elm et al., 2007). All continuous variables will be summarised using the following descriptive statistics:

- N (size of analysis population without missing values)

- Arithmetic mean
- Standard deviation (SD)
- Median
- 25th percentile value (P25), 75th percentile value (P75) and interquartile range (IQR)
- Minimum and maximum (where relevant)

The proportion/percentage of observed levels will be reported for all categorical measures. Where appropriate, corresponding exact 95% confidence intervals (CIs) for proportions will be included.

Reporting Conventions

P-values ≥ 0.001 and ≤ 0.999 will be reported to 3 decimal places; *p*-values less than 0.001 will be reported as "< 0.001". The mean, standard deviation, median, IQR and other statistics will be reported to 3 decimal places. Minimum and maximum values will use the same number of decimal places as the original data. Proportions and percentages will be presented as 2 decimal places; values greater than zero but < 0.01 will be presented as "< 0.01", and values greater than 99.99% but less than 100% will be reported as "> 99.99%". Estimated parameters, not on the same scale as raw observations (e.g. regression coefficients) will be reported to 3 significant figures.

Missing Data

While every effort will be made to follow-up enrolled participants, there may be loss-to-follow-up. The primary analyses, including examination of longitudinal resistome evolution, will only include participants with complete follow-up data. Analyses for all secondary outcomes will include all participants. Missing data will be fully described.

Technical Details

The R environment for statistical computing (v4.2.2 or later) will be used for all analyses. All analysis code will be made publicly available on GitHub.

Participant Characteristics

We will list the participant's characteristics by the observation group within which they were enrolled, see Tables 2 to 5.

Table 2. (MOCK TABLE) Study participant characteristics

	Total (N = 350)	Healthy community (N = 83)	Community pneumonia (N = 92)	1st admission with pneumonia (N = 92)	Re-admission with pneumonia (N = 83)
Age (months)					
Mean (SD)	N (SD)	N (SD)	N (SD)	N (SD)	N (SD)
Sex					
Female	N (%)	N (%)	N (%)	N (%)	N (%)
Male	N (%)	N (%)	N (%)	N (%)	N (%)
Co-morbidities					
Present	N (%)	N (%)	N (%)	N (%)	N (%)
Mode of delivery					
NVD	N (%)	N (%)	N (%)	N (%)	N (%)
C-section	N (%)	N (%)	N (%)	N (%)	N (%)
Exclusively breastfed 1st 3 months of life					
Yes	N (%)	N (%)	N (%)	N (%)	N (%)
Indoor household pollutant exposure					
Yes	N (%)	N (%)	N (%)	N (%)	N (%)
Received PCV?					
1 dose	N (%)	N (%)	N (%)	N (%)	N (%)
2 doses	N (%)	N (%)	N (%)	N (%)	N (%)
3 doses	N (%)	N (%)	N (%)	N (%)	N (%)
No	N (%)	N (%)	N (%)	N (%)	N (%)
Received HiB vaccine?					
1 dose	N (%)	N (%)	N (%)	N (%)	N (%)
2 doses	N (%)	N (%)	N (%)	N (%)	N (%)
3 doses	N (%)	N (%)	N (%)	N (%)	N (%)
No	N (%)	N (%)	N (%)	N (%)	N (%)
MUAC (cm)					
Mean (SD)	N (SD)	N (SD)	N (SD)	N (SD)	N (SD)
NPS viral PCR result					
Influenza A	N (%)	N (%)	N (%)	N (%)	N (%)
Influenza B	N (%)	N (%)	N (%)	N (%)	N (%)
RSV subtype A	N (%)	N (%)	N (%)	N (%)	N (%)
RSV subtype B	N (%)	N (%)	N (%)	N (%)	N (%)
Negative	N (%)	N (%)	N (%)	N (%)	N (%)
Healthcare seeking behaviour					
Time taken to seek healthcare (hours)					
< 1	N (%)	N (%)	N (%)	N (%)	N (%)
1-6	N (%)	N (%)	N (%)	N (%)	N (%)
7-12	N (%)	N (%)	N (%)	N (%)	N (%)
> 12 hours	N (%)	N (%)	N (%)	N (%)	N (%)
Parent educational level					
Primary school	N (%)	N (%)	N (%)	N (%)	N (%)
Secondary school	N (%)	N (%)	N (%)	N (%)	N (%)
College/University	N (%)	N (%)	N (%)	N (%)	N (%)
Unknown	N (%)	N (%)	N (%)	N (%)	N (%)
Prior antibiotic exposure					
Antibiotic exposure in preceding 7 days					
Yes	N (%)	N (%)	N (%)	N (%)	N (%)
Duration of preceding antibiotic exposure (days)					
Mean (SD)	N (SD)	N (SD)	N (SD)	N (SD)	N (SD)
Class of preceding antibiotic exposure					
Beta-lactam	N (%)	N (%)	N (%)	N (%)	N (%)
Macrolide	N (%)	N (%)	N (%)	N (%)	N (%)
Tetracycline	N (%)	N (%)	N (%)	N (%)	N (%)
Phenicol	N (%)	N (%)	N (%)	N (%)	N (%)
Aminoglycoside	N (%)	N (%)	N (%)	N (%)	N (%)
Other	N (%)	N (%)	N (%)	N (%)	N (%)

Table 3. (MOCK TABLE) Hospitalised participant characteristics

	Total (N = 175)	1st admission with pneumonia (N = 92)	Re-admission with pneumonia (N = 83)
Haemoglobin (g/dL)			
Mean (SD)	N (SD)	N (SD)	N (SD)
Malaria point-of-care test positive			
Yes	N (%)	N (%)	N (%)
Time since last hospital admission (days)			
Mean (SD)	N (SD)	N (SD)	N (SD)
Markers of pneumonia severity			
Central cyanosis	N (%)	N (%)	N (%)
Oxygen saturations < 90%	N (%)	N (%)	N (%)
Inability to feed	N (%)	N (%)	N (%)
Vomiting	N (%)	N (%)	N (%)
Convulsions	N (%)	N (%)	N (%)
Lethargy/impaired conscious level	N (%)	N (%)	N (%)
Duration of hospital admission (days)			
Mean (SD)	N (SD)	N (SD)	N (SD)
Duration of inpatient bed-sharing (days)			
Mean (SD)	N (SD)	N (SD)	N (SD)
Number of inpatients bed-sharing			
Mean (SD)	N (SD)	N (SD)	N (SD)

Table 4. (MOCK TABLE) Antibiotic prescriptions for pneumonia treatment

	Total (N = 267)	Community pneumonia (N = 92)	1st admission with pneumonia (N = 92)	Re-admission with pneumonia (N = 83)
Total duration of antibiotic prescribed for pneumonia (days)				
Mean (SD)	N (SD)	N (SD)	N (SD)	N (SD)
Class of outpatient antibiotic prescription				
Beta-lactam	N (%)	N (%)	N (%)	N (%)
Macrolide	N (%)	N (%)	N (%)	N (%)
Tetracycline	N (%)	N (%)	N (%)	N (%)
Chloramphenicol	N (%)	N (%)	N (%)	N (%)
Aminoglycoside	N (%)	N (%)	N (%)	N (%)
Other	N (%)	N (%)	N (%)	N (%)
Duration of outpatient antibiotic prescription (days)				
Mean (SD)	N (SD)	N (SD)	N (SD)	N (SD)
Class of inpatient antibiotic prescription				
Beta-lactam	N (%)	N (%)	N (%)	N (%)
Macrolide	N (%)	N (%)	N (%)	N (%)
Tetracycline	N (%)	N (%)	N (%)	N (%)
Phenicol	N (%)	N (%)	N (%)	N (%)
Aminoglycoside	N (%)	N (%)	N (%)	N (%)
Other	N (%)	N (%)	N (%)	N (%)
Duration of inpatient antibiotic prescription (days)				
Mean (SD)	N (SD)	N (SD)	N (SD)	N (SD)

Table 5. (MOCK TABLE) Participant characteristics at 3-month follow up

	Total (N = 267)	Community pneumonia (N = 92)	1st admission with pneumonia (N = 92)
3-month mortality (count)			
Yes	N (%)	N (%)	N (%)
Re-presentation to hospital or healthcare centre with pneumonia in past 3-months?			
Yes	N (%)	N (%)	N (%)
Re-admission to hospital with pneumonia in past 3-months?			
Yes	N (%)	N (%)	N (%)
Duration of re-hospitalisation (days)			
Mean (SD)	N (SD)	N (SD)	N (SD)
Additional outpatient antibiotic prescription in past 3-months for pneumonia			
Yes	N (%)	N (%)	N (%)
Number of additional outpatient antibiotic prescription in past 3-months for pneumonia			
Mean (SD)	N (SD)	N (SD)	N (SD)
Class of additional outpatient antibiotic prescription			
Beta-lactam	N (%)	N (%)	N (%)
Macrolide	N (%)	N (%)	N (%)
Tetracycline	N (%)	N (%)	N (%)
Phenicol	N (%)	N (%)	N (%)
Aminoglycoside	N (%)	N (%)	N (%)
Other	N (%)	N (%)	N (%)
Duration of additional outpatient antibiotic prescription (days)			
Mean (SD)	N (SD)	N (SD)	N (SD)

Analysis of Primary Endpoints

The Bray-Curtis Dissimilarity Index will be calculated to quantify the beta diversity, or compositional (dis)similarity between the NPR of the four study groups. This will be used to generate Non-metric Multidimensional Scaling (NMDS) plots to visualise the (dis)similarity between the NPR composition of the different groups (see Figure 3); these plots map items based on distance, to represent the pairwise dissimilarity between objects in a two-dimensional space. The Bray-Curtis Dissimilarity Index will be compared between groups using a permutational multivariate analysis of variance (PERMANOVA).

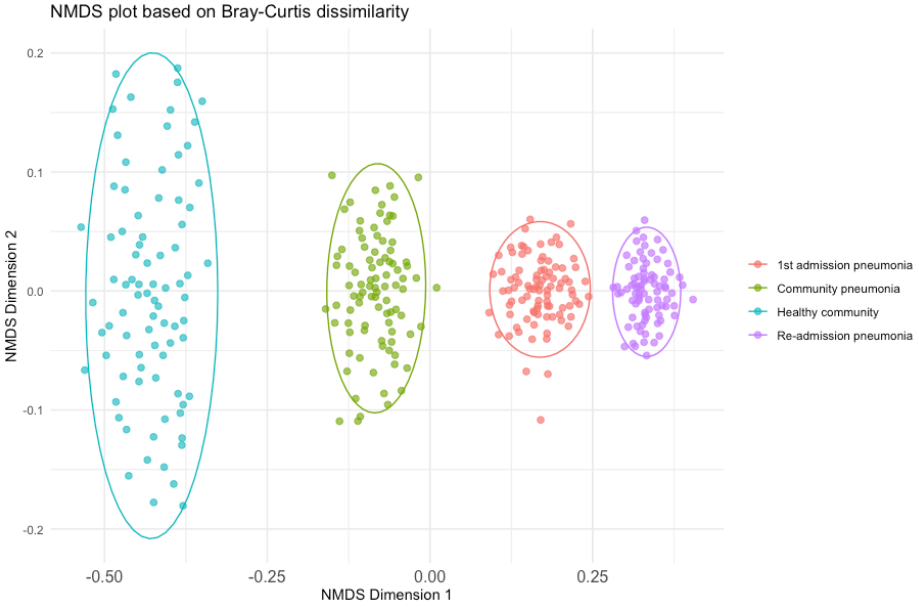

Figure 3. (MOCK FIGURE) NMDS plot comparing Bray Curtis dissimilarity index between the four study groups.

The relative abundance of ARG classes will be compared between groups visually using stacked bar plots (see Figure 4) and/or heatmaps with antibiotic usage data (see Figure 5), and statistically using the Kruskal-Wallis test with Dunn's post-hoc pairwise comparison, or Wilcoxon Signed-Rank test for paired samples (from the same participant at different timepoints), to compare the median values between groups with false discovery rate correction.

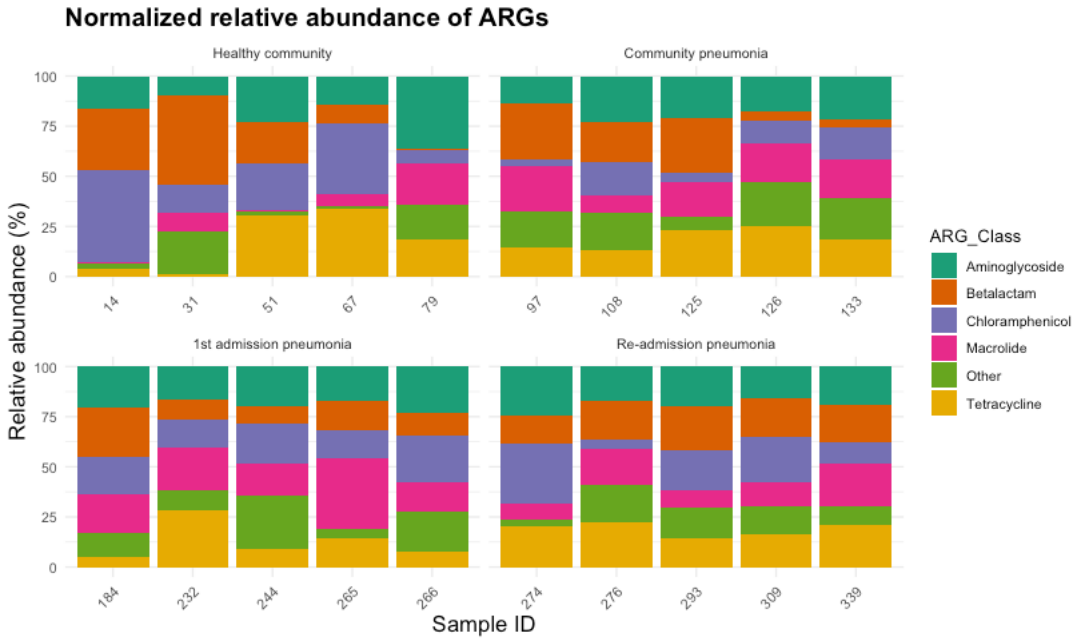

Figure 4. (MOCK FIGURE) Bar chart comparing the relative abundance of ARG classes between samples in healthy community, community pneumonia, 1st admission pneumonia and re-admission pneumonia groups.

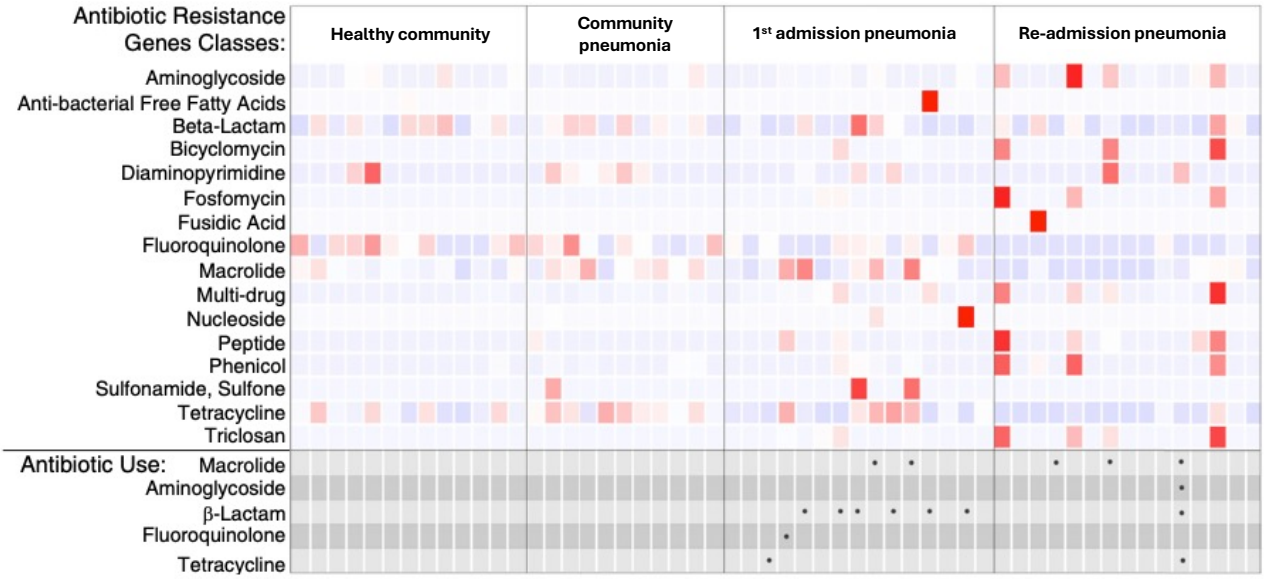

Figure 5. (MOCK FIGURE) Heatmap comparing the relative abundance of ARG classes between samples in healthy community, community pneumonia, 1st admission pneumonia and re-admission pneumonia groups, with antibiotic consumption data recorded below.

Primary Endpoint Expectations:

1. There will be a difference in the overall composition of the NPR (beta diversity) between all the groups. There will be an increase in the relative abundance of ARG classes that produce AMR to first line antibiotics used to treat pneumonia within the NPR between healthy community and all pneumonia participants; hospitalised groups will have the highest relative abundance of these ARG classes.
2. There will be a difference in the overall composition of the NPR (beta diversity) between recruitment and discharge time-points in hospital pneumonia participants, but no difference between discharge and 3-month follow up time-points. There will be an increase in the relative abundance of ARG classes that produce AMR to first line antibiotics used to treat pneumonia within the NPR between recruitment and discharge time-points in hospital pneumonia participants. There will be no difference in the relative abundance of these ARG classes between discharge and 3-month follow up time-points.
3. There will be no difference in the overall composition of the NPR (beta diversity) between recruitment and 3-month follow up in community pneumonia participants. There will be no difference in the relative abundance of ARG classes that produce AMR to first line antibiotics used to treat pneumonia within the NPR between recruitment and 3-month follow up in community pneumonia participants.

The statistical methods and visual comparisons used to test these hypotheses (primary study outcomes 1 to 3) are summarised in Table 7, below.

Analysis of Secondary Endpoints

The relative abundance of individual ARGs will be compared between groups as detailed for the ARG classes, above. The median Shannon Diversity index will be calculated to estimate the alpha diversity of the NPR in each study group; this will be compared between study groups visually using box-plots (see Figure 6), and statistically using the Kruskal-Wallis test with Dunn's post-hoc pairwise comparison, or Wilcoxon Signed-Rank test for paired samples (different time-points from the same participant) with false discovery rate correction.

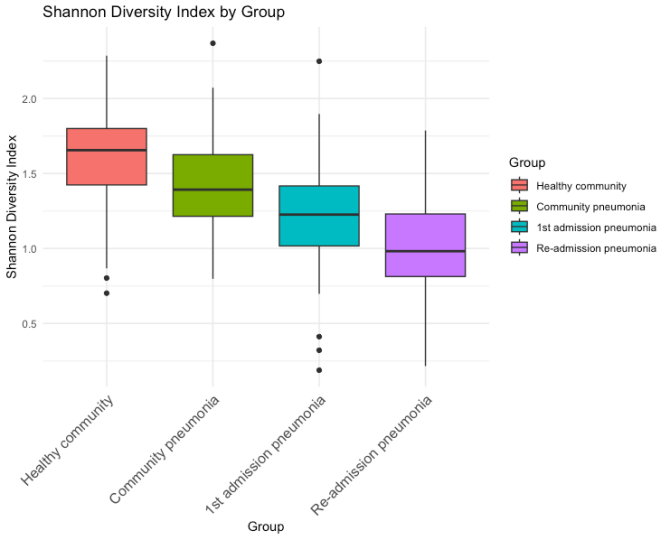

Figure 6. (MOCK FIGURE) Boxplot comparing median Shannon diversity index for bacterial species diversity across the four study groups.

Secondary Endpoint Expectations:

1. There will be a difference in the alpha diversity of the NPR between the groups; healthy community participants will have the least diverse NPR and hospitalised groups will have the most diverse NPR. There will be an increase in the relative abundance of individual ARGs that produce AMR to first line antibiotics used to treat pneumonia within the NPR between healthy community and all pneumonia participants; hospitalised groups will have the highest relative abundance of these ARGs.

There will be a difference in NPR composition between healthy community and hospital readmission groups at enrolment, and community pneumonia and hospital pneumonia groups at the 3-month study end-point based on hierarchical clustering of Bray-Curtis dissimilarity indices.

To test this hypothesis, NPR composition will be classified into hierarchical clusters based on Bray-Curtis dissimilarities of individual ARGs, using Ward's method of hierarchical clustering to minimise within-cluster variance. Hierarchical clusters will be visualised using dendrograms to determine the optimal number of clusters to differentiate NPR composition, as illustrated below in Figure 7. The association of the compositional clusters with clinical metadata will also be examined by applying NMDS to the Bray-Curtis dissimilarities and fitting clinical covariates (pneumonia severity, RSV co-infection) as vectors using the envfit function from the vegan package in R with 999 permutations to assess significance, as illustrated below in Figure 8.

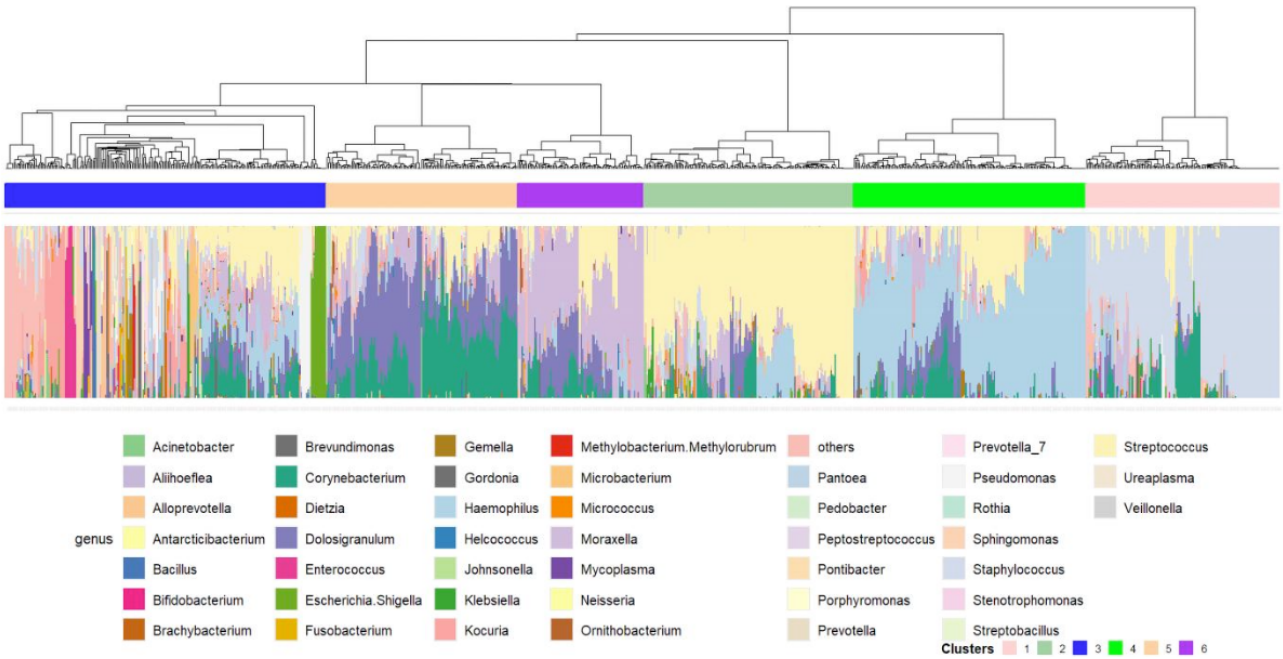

Figure 7. An example of hierarchical clustering of genus-level relative abundance profiles to reveal distinct NPM clusters based on Bray-Curtis dissimilarity and Ward's linkage (Makori et al., 2026).

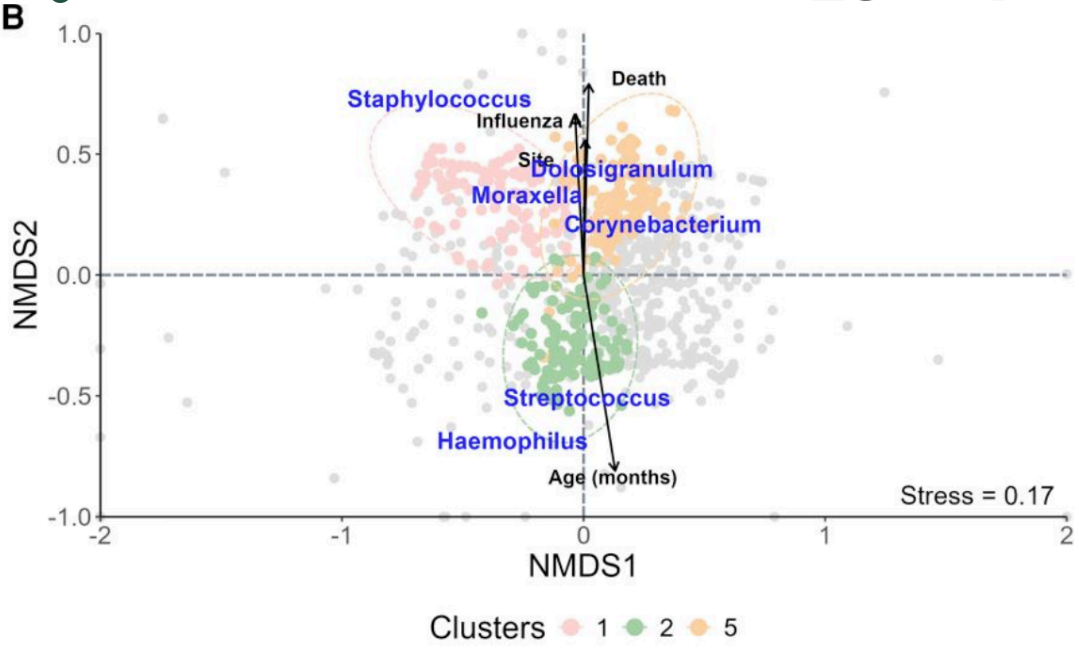

Figure 8. An example of a NMDS of Bray-Curtis dissimilarities (based on bacterial genus) overlaid with microbial cluster assignments and arrows representing clinical outcomes significantly associated with gradients in the NPM structure (Makori et al., 2026).

2. There will be no difference in the alpha diversity of the NPR between discharge and 3-month follow up in hospitalised pneumonia participants; the NPR will be more diverse at discharge than recruitment. There will be an increase in the relative abundance of individual ARGs that produce AMR to first line antibiotics used to treat pneumonia within the NPR between recruitment and discharge time-points in hospital pneumonia participants. There will be no difference in the relative abundance of these ARGs between discharge and 3-month follow up time-points.
3. There will be no difference in the alpha diversity of the NPR between recruitment and 3-month follow up in community pneumonia participants. There will be no difference in the relative abundance of individual ARGs that produce AMR to first line antibiotics used to treat pneumonia within the NPR between recruitment and 3-month follow up in community pneumonia participants.

The statistical methods and visual comparisons used to test these hypotheses (secondary study endpoints 1 to 3) are summarised in Table 8, below, except for the hierarchical clustering (detailed above).

4. a) There will be a difference in the alpha diversity of the NPR between the groups; healthy community and community pneumonia participants will have a less diverse NPR than hospital participants at discharge. Hospital readmission participants will have the most diverse NPR.
   b) There will be a difference in the overall composition of the NPR (beta diversity) between all the groups.
   c)/d) There will be an increase in the relative abundance of ARG classes and individual ARGs that produce AMR to first line antibiotics used to treat pneumonia within the NPR between healthy community and all pneumonia participants; hospital

readmission participants will have the highest relative abundance of these ARG classes.

5. a) There will be a difference in the alpha diversity of the NPR between the groups; community pneumonia participants at enrolment will have a more diverse NPR than healthy community participants. Hospital readmission participants will have the most diverse NPR.
   b) There will be a difference in the overall composition of the NPR (beta diversity) between all the groups.
   c)/d) There will be an increase in the relative abundance of ARG classes and individual ARGs that produce AMR to first line antibiotics used to treat pneumonia within the NPR between healthy community and all pneumonia participants; hospitalised participants will have a higher relative abundance of these ARG classes than community pneumonia participants at enrolment.
6. a) There will be a difference in the alpha diversity of the NPR between the groups; healthy community participants will have the least diverse NPR. Hospital readmission participants will have the most diverse NPR. Community pneumonia participants at enrolment will have a less diverse NPR than hospital participants at discharge.
   b) There will be a difference in the overall composition of the NPR (beta diversity) between all the groups.
   c)/d) There will be an increase in the relative abundance of ARG classes and individual ARGs that produce AMR to first line antibiotics used to treat pneumonia within the NPR between healthy community and all pneumonia participants; hospitalised participants will have a higher relative abundance of these ARG classes than community pneumonia participants at enrolment.
7. a) There will be a difference in the alpha diversity of the NPM between the groups; healthy community participants will have the most diverse NPM. Hospitalised participants will have the least diverse NPM.
   b) There will be a difference in the overall composition of the NPM (beta diversity) between all the groups.
   c) There will be a difference in the relative abundance of bacterial taxa within the NPM between healthy community and all pneumonia participants. Healthy community participants will have the highest relative abundance of commensal bacterial taxa. Hospital readmission participants will have the highest relative abundance of pathogenic bacterial taxa.
8. a) There will be no difference in the alpha diversity of the NPM between discharge and 3-month follow up in hospitalised pneumonia participants; the NPM will be less diverse at discharge than recruitment.
   b) There will be a difference in the overall composition of the NPM (beta diversity) between recruitment and discharge time-points in hospital pneumonia participants, but no difference between discharge and 3-month follow up time-points.
   c) There will be a difference in the relative abundance of bacterial taxa within the NPM between recruitment and discharge time-points in hospital pneumonia participants, but no difference between discharge and 3-month follow up time-points.
9. a) There will be no difference in the alpha diversity of the NPM between recruitment and 3-month follow up in community pneumonia participants.

b) There will be no difference in the overall composition of the NPM (beta diversity) between recruitment and 3-month follow up in community pneumonia participants.
c) There will be no difference in the relative abundance of bacterial taxa within the NPM between recruitment and 3-month follow up in community pneumonia participants.

The statistical methods and visual comparisons used to test these hypotheses (secondary study endpoints 5 to 10) are summarised in Table 8, below.

10. The hospital pneumonia group will have a higher rate of repeat health clinic attendances, antibiotic prescriptions and hospitalisation at the 3-month study end-point than the community pneumonia group.

The rate of repeat health clinic attendance, antibiotic prescription, hospitalisation and mortality within 3 months will be compared between the community pneumonia and hospitalised pneumonia groups visually using a bar chart, and statistically using a Wilcoxon Rank Sum test.

Exploratory Analyses Expectations:

11. There will be a difference in NPR composition between healthy community and hospital readmission groups at enrolment, and community pneumonia and hospital pneumonia groups at the 3-month study end-point based on hierarchical clustering of Bray-Curtis dissimilarity indices.

To test this hypothesis, NPR composition will be classified into hierarchical clusters based on Bray-Curtis dissimilarities of ARG classes, using Ward's method of hierarchical clustering to minimise within-cluster variance. Hierarchical clusters will be visualised using dendrograms to determine the optimal number of clusters to differentiate NPR composition (Figure 7). As in secondary endpoint analysis 1, the association of the compositional clusters with clinical metadata will also be examined by applying NMDS to the Bray-Curtis dissimilarities and fitting clinical covariates (pneumonia severity, RSV co-infection) as vectors using the envfit function from the vegan package in R with 999 permutations to assess significance (Figure 8).

12. There will be an association between NPR composition, as defined by hierarchical clustering of Bray-Curtis dissimilarities based on ARG classes, and recurrent hospitalisation with pneumonia.

This hypothesis will be tested using the hierarchical clusters defined above (exploratory analysis 11) for NPS samples from the hospital readmission group at enrolment and hospital pneumonia group at the 3-month study end-point. Compositional cluster membership will be used as the exposure in a logistic regression model, where the outcomes of interest are more than 1 hospital admission with pneumonia within 3 months (cases) versus no hospital admission within 3 months of discharge after hospitalisation with pneumonia (controls). Potential confounders and sources of interaction will be identified by a directed acyclic graph, and included in the model if there is evidence of confounding or effect modification (based on a difference between adjusted and crude effect estimates).

13. There will be an association between NPR composition at discharge and recurrent hospitalisation within 3 months with pneumonia among hospitalised participants.

This hypothesis will be tested using a logistic regression model to test for association between NPR composition at discharge from hospital in the hospitalised participants (the exposure), and re-hospitalisation with pneumonia within 3-months (the outcome). NPR composition at discharge from hospital will be defined by:

a) ARG richness and evenness (alpha diversity) as measured by the Shannon diversity index of each sample.
b) Hierarchical clustering of Bray-Curtis dissimilarities based on ARGs, using Ward's method of hierarchical clustering to minimise within-cluster variance; the exposure will be defined by cluster membership.
c) The relative abundance of ARGs to beta-lactam antibiotics, as calculated by RPKM.

The study is not powered for this outcome, so these will be reported as unadjusted odds ratios, with 95% confidence intervals. As above, hierarchical clusters will be visualised using dendrograms to determine the optimal number of clusters to differentiate NPR composition. The association of the compositional clusters with clinical metadata will also be examined by applying NMDS to the Bray-Curtis dissimilarities and fitting clinical covariates (pneumonia severity, RSV co-infection) as vectors.

14. There will be a difference in the alpha diversity, beta diversity and relative abundance of bacterial taxa in the NPM at recruitment when compared between all hospitalised participants with pneumonia or severe pneumonia:
    a) Species richness and evenness (alpha diversity) of the NPM
    b) Overall taxonomic composition (beta diversity) of the NPM
    c) Relative abundance of bacterial taxa in the NPM

Severe pneumonia will be defined as described in outcome 11, using the WHO classification as cough or difficulty breathing with at least one of the following:

- Oxygen saturations < 90% or central cyanosis
- Severe respiratory distress
- General danger signs (inability to feed, lethargy, reduced level of consciousness or convulsions) (World Health Organisation, 2013)

15. There will be a difference in the alpha diversity, beta diversity, relative abundance of individual ARGs and ARG classes in the NPR when compared between all hospitalised participants with pneumonia or severe pneumonia.
16. There will be a difference in the alpha diversity, beta diversity and relative abundance of bacterial taxa in the NPM at recruitment when compared between all hospitalised participants with viral pathogens detected at recruitment.

17. There will be a difference in the alpha diversity, beta diversity, relative abundance of individual ARGs or ARG classes in the NPR when compared between all hospitalised participants with viral pathogens detected at recruitment.

The statistical methods and visual comparisons used to test these hypotheses (exploratory analyses 14 to 17) are summarised in Table 8, below.

18. a) Participants with a longer duration of hospitalisation will have a more diverse NPR.
    b) Participants with a longer duration of hospitalisation will have a higher relative abundance of ARG classes and individual ARGs that produce AMR to first line antibiotics used to treat pneumonia within the NPR.

This hypothesis will be tested using simple linear regression models to test for association between NPR composition (relative abundance of specific ARG classes measured by RPKM and change in beta-diversity using within participant recruitment-to-discharge Bray-Curtis dissimilarity indices) and duration of hospitalisation. All analyses of duration of hospitalisation and bed-space sharing will be adjusted for pneumonia severity and antibiotic duration.

19. a) Participants from shared bed spaces will have a more diverse NPR (alpha diversity) when compared to participants that did not share a bed-space.
    b) Participants from shared bed-spaces will have a different overall NPR composition (beta diversity) when compared to participants that did not share a bed-space.
    c) Participants from shared bed-spaces will have a higher relative abundance of ARG classes and individual ARGs associated with environmental gram-negative organisms when compared to participants that did not share a bed-space.

20. There will be a difference in the overall composition of the NPR (beta diversity) between participants with and without antibiotic use prior to recruitment, as detected by the urine antibiotic bioassay, within each study group. Participants with prior antibiotic use detected will have lower alpha diversity and a higher relative abundance of ARG classes used as first-line for antibiotic therapy for pneumonia.

The statistical methods and visual comparisons used to test these hypotheses are summarised in Table 8, below.

21. There will be an increase in the relative abundance of MGE between enrolment and discharge in the NPR of hospitalised participants.

This hypothesis will be tested by visually comparing the relative abundance of MGEs between enrolment and discharge time points, using stacked bar plots of RPKM, and statistically by comparing the median relative abundance using the Wilcoxon Signed-Rank test for paired samples with false discovery rate correction.

Table 7. Illustrates for each primary endpoint, the sample groups being compared, hypotheses being tested, methods of data visualisation and statistical tests being used.

Primary Study Endpoints	Group 1	Group 2	Group 3	Group 4	Hypothesis	Statistical analysis plan
1	Healthy community	Community pneumonia, NPS at 3-month follow up	Hospitalised pneumonia, NPS at 3-month follow up	Re-hospitalised pneumonia, NPS at enrolment	There will be a difference in the following measures of NPR composition between groups: a) Beta diversity b) Relative abundance of ARG classes	a) Beta diversity • Visually compare the Bray-Curtis Dissimilarity Index between groups using NMDS plots. • Statistically compare the Bray-Curtis Dissimilarity Index between groups using PERMANOVA (Endpoint 1), or paired PERMANOVA with permutations restricted within participants for comparisons between time points (Endpoints 2 and 3). b) Relative abundance of ARG classes • Visually compare the relative abundance between groups using stacked bar plots of RPKM. • Statistically compare the median relative abundance between all four groups using the Kruskal-Wallis test with Dunn's post-hoc pairwise comparisons and false discovery rate correction (to account for one test performed for each class of ARG). • Statistically compare the median relative abundance between timepoints within participants using the Wilcoxon Signed-Rank test with false discovery rate correction (Endpoints 2 and 3). • Statistically compare the median relative abundance of ARG classes between community pneumonia at enrolment and 3-month follow up, using a two-one-sided-tests (TOST) approach for equivalence.
2	Hospitalised pneumonia, NPS at enrolment	Hospitalised pneumonia, NPS at discharge	Hospitalised pneumonia, NPS at 3-month follow up	N/A	There will be a difference in measures a) and b) of NPR composition between groups 1 and 2, and groups 1 and 3.	
3	Community pneumonia, NPS at enrolment	Community pneumonia, NPS at 3-month follow up	N/A	N/A	There will be ***no difference*** in the measures a) and b) of NPR composition between groups 1 and 2.	

Table 8. Illustrates the sample groups being compared, hypotheses being tested, methods of data visualisation and statistical tests being used for secondary endpoints and exploratory analyses (where this is not detailed in the main body of text).

Secondary Study Endpoints	Group 1	Group 2	Group 3	Group 4	Hypothesis	Statistical analysis plan
1	Healthy community	Community pneumonia, NPS at 3-month follow up	Hospitalised pneumonia, NPS at 3-month follow up	Re-hospitalised readmission pneumonia, NPS at enrolment	There will be a difference in the following measures of NPR composition between groups: a) Alpha diversity b) Relative abundance of individual ARGs	a) Alpha diversity • Visually compare the median Shannon Diversity Index of the NPR between groups using a boxplot. • Statistically compare the median Shannon Diversity Index of the NPR between all four groups using a Kruskal Wallis test with Dunn's post-hoc pairwise comparison (Endpoint 1). • Statistically compare the median Shannon Diversity Index of the NPR between timepoints within participants using the Wilcoxon Signed-Rank test for paired samples (Endpoints 2 and 3). * Statistically compare the alpha diversity between timepoints within participants, where no statistically significant difference is expected, using two-one-sided-tests (TOST) for equivalence. b) Relative abundance of individual ARGs • Visually compare the relative abundance between groups using stacked bar plots of RPKM. • Statistically compare the median relative abundance between all four groups using the Kruskal-Wallis test with Dunn's post-hoc pairwise comparison and false discovery rate correction (Endpoint 1). • Statistically compare the median relative abundance between groups of two using the Wilcoxon Signed-Rank test for paired samples with false discovery rate correction (Endpoint 2 and 3). * Statistically compare the median relative abundance of ARGs between timepoints within participants, where no statistically significant difference is expected, using two-one-sided-tests (TOST) for equivalence.
2	Hospitalised pneumonia, NPS at enrolment	Hospitalised pneumonia, NPS at discharge	Hospitalised pneumonia, NPS at 3-month follow up	N/A	There will be a difference in measures a) to b) of NPR composition between groups 1 and 2, and groups 1 and 3, but ***no*** statistically significant difference between groups 2 and 3.*	
3	Community pneumonia, NPS at enrolment	Community pneumonia, NPS at 3-month follow up	N/A	N/A	There will be ***no difference*** in the following measures a) to b) of NPR composition between groups 1 and 2.*	
4	Healthy community	Community pneumonia, NPS at 3-month follow up	Hospitalised pneumonia, NPS at discharge	Re-hospitalised pneumonia, NPS at enrolment	There will be a difference in the following measures of NPR composition between groups: a) Alpha diversity b) Beta diversity c) Relative abundance of ARG classes d) Relative abundance of individual ARGs	a) Alpha diversity • Visually compare the median Shannon Diversity Index of the NPR or NPM between groups using a boxplot. • Statistically compare the median Shannon Diversity Index of the NPR or NPM between all four groups using a Kruskal Wallis test with Dunn's post-hoc pairwise comparison and false discovery rate correction (Endpoints 4 to 7, 14 to 17, 19 and 20). • Statistically compare the median Shannon Diversity Index of the NPM between groups of two using the Wilcoxon Signed-Rank test for paired samples with false discovery rate correction (Endpoints 8 and 9). * Statistically compare the median Shannon Diversity Index of the NPR or NPM, between groups where no statistically significant difference is expected using two-one-sided-tests (TOST) for equivalence.
5	Healthy community	Community pneumonia, NPS at enrolment	Hospitalised pneumonia, NPS at 3-month follow up	Re-hospitalised pneumonia, NPS at enrolment		
6	Healthy community	Community pneumonia, NPS at	Hospital pneumonia, NPS at	Re-hospitalised pneumonia,		

		Enrolment	discharge	NPS at enrolment	
19	Hospital pneumonia, NPS at discharge				
	Shared bed-space	No shared bed-space	N/A	N/A	
20	Healthy community at enrolment				
	Antibiotic use detected	No antibiotic use detected	N/A	N/A	
	Community pneumonia, NPS at enrolment				
	Antibiotic use detected	No antibiotic use detected	N/A	N/A	
	Hospital pneumonia, NPS at enrolment				
	Antibiotic use detected	No antibiotic use detected	N/A	N/A	
	Re-hospitalised pneumonia, NPS at enrolment				
	Antibiotic use detected	No antibiotic use detected	N/A	N/A	
7	Healthy community	Community pneumonia, NPS at 3-month follow up	Hospital pneumonia, NPS at 3-month follow up	Re-hospitalised pneumonia, NPS at enrolment	There will be a difference in the following measures of NPM composition between groups: a) Alpha diversity b) Beta diversity c) Relative abundance of bacterial taxa
8	Hospital pneumonia, NPS at enrolment	Hospital pneumonia, NPS at discharge	Hospital pneumonia, NPS at 3-month follow up	N/A	There will be a difference in measures a) to c) of NPM composition between groups 1 and 2, and groups 1 and 3, but ***no*** statistically significant difference between groups 2 and 3.*
9	Community pneumonia, NPS at enrolment	Community pneumonia, NPS at 3-month follow up	N/A	N/A	There will be ***no difference*** in the measures a) to c) of NPM composition between groups.*
14/15	Hospitalised pneumonia and re-hospitalised pneumonia participants, NPS at enrolment				There will be a difference in the following measures of NPM and NPR composition between groups: a) Alpha diversity
	Pneumonia	Severe pneumonia	N/A	N/A	

b) Beta diversity
- Visually compare the Bray-Curtis Dissimilarity Index between groups using NMDS plots.
- Statistically compare the Bray-Curtis Dissimilarity Index between groups using PERMANOVA (Endpoints 4 to 7, 14 to 17, 19 and 20), or paired PERMANOVA for timepoints within participants (Endpoints 8 and 9).

c/d) Relative abundance of ARG classes/individual ARGs for the NPR; relative abundance of bacterial taxa for the NPM
- Visually compare the relative abundance between groups using stacked bar plots of RPKM.
- Statistically compare the median relative abundance between all four groups using the Kruskal-Wallis test with Dunn's post-hoc pairwise comparison and false discovery rate correction (Endpoints 4 to 7, 14 to 17, 19 and 20).
- Statistically compare the median relative abundance between groups of two using the Wilcoxon Signed-Rank test for paired samples with false discovery rate correction (Endpoints 8 and 9).

*Statistically compare the median relative abundance of ARG classes/individual ARGs for the NPR, or relative abundance of bacterial taxa for the NPM, between groups where no statistically significant difference is expected using two-one-sided-tests (TOST) for equivalence.

16/17	Influenza A	Influenza B	RSV	No viral pathogen detected	b) Beta diversity c/d) Relative abundance of bacterial taxa/ARG classes/individual ARGs	

References

Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *44*(2), 139–177.

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., & Weyrich, L. S. (2019). Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends in Microbiology*, *27*(2), 105–117. https://doi.org/10.1016/j.tim.2018.11.003

Ferdous, T., Jiang, L., Dinu, I., Groizeleau, J., Kozyrskyj, A. L., Greenwood, C. M. T., & Arrieta, M.-C. (2022). The rise to power of the microbiome: Power and sample size calculation for microbiome studies. *Mucosal Immunology*, *15*(6), 1060–1070. https://doi.org/10.1038/s41385-022-00548-1

Hou, J., Song, Y., Leung, A. S. Y., Tang, M. F., Shi, M., Wang, E. Y., Tsun, J. G. S., Chan, R. W. Y., Wong, G. W. K., Tsui, S. K.-W., & Leung, T. F. (2022). Temporal Dynamics of the Nasopharyngeal Microbiome and its Relationship with Childhood Asthma Exacerbation. *Microbiology Spectrum*, *10*(3), e0012922. https://doi.org/10.1128/spectrum.00129-22

Limwado, G. D., Aron, M. B., Mpinga, K., Phiri, H., Chibvunde, S., Banda, C., Ndarama, E., Walyaro, C., & Connolly, E. (2024). Prevalence of antibiotic self-medication and knowledge of antimicrobial resistance among community members in Neno District rural Malawi: A cross-sectional study. *IJID Regions*, *13*, 100444. https://doi.org/https://doi.org/10.1016/j.ijregi.2024.100444

Makori, T. O., Gicheru, E. T., Mburu, M. W., Sada, M. S., Nyawa, O., Mutunga, M., Lewa, C., Cheruiyot, R., Kiguli, S., Olupot-Olupot, P., Muhindo, R., Mogaka, C., Williams, T. N., Agoti, C. N., Maitland, K., & Sande, C. J. (2026). Nasopharyngeal Microbiome Composition and its Clinical Correlates in Children Hospitalized with Severe Pneumonia in East Africa. *The Journal of Infectious Diseases*, *234*(3), 475–483. https://doi.org/10.1093/infdis/jiag093

Tonkin-Hill, G., Ling, C., Chaguza, C., Salter, S. J., Hinfonthong, P., Nikolaou, E., Tate, N., Pastusiak, A., Turner, C., Chewapreecha, C., Frost, S. D. W., Corander, J., Croucher, N. J., Turner, P., & Bentley, S. D. (2022). Pneumococcal within-host diversity during colonization, transmission and treatment. *Nature Microbiology*, *7*(11), 1791–1804. https://doi.org/10.1038/s41564-022-01238-1

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., & Initiative, for the S. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *PLOS Medicine*, *4*(10), e296-.

World Health Organisation. (2013). Cough or Difficulty in Breathing. In *Pocket book of hospital care for children: Guidelines for the management of common childhood illnesses 2nd edition* (2nd edn, pp. 75–103). World Health Organisation. https://iris.who.int/server/api/core/bitstreams/8f110da0-22e6-4ef1-90e4-c9f1b7daa363/content

Zhang, Z., Zhang, Q., Wang, T., Xu, N., Lu, T., Hong, W., Penuelas, J., Gillings, M., Wang, M., Gao, W., & Qian, H. (2022). Assessment of global health risk of antibiotic resistance genes. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-29283-8