

# Transmission dynamics and control of multidrug-resistant *Klebsiella pneumoniae* in neonates in a developing country

Thomas Crellen<sup>1,2\*</sup>, Paul Turner<sup>1,2,3</sup>, Sreymom Pol<sup>1,3</sup>, Stephen Baker<sup>2,4</sup>, To Nguyen Thi Nguyen<sup>4</sup>, Nicole Stoesser<sup>2</sup>, Nicholas P.J. Day<sup>1,2</sup>, Claudia Turner<sup>1,2,3</sup>, Ben S. Cooper<sup>1,2\*</sup>

\*For correspondence:

thomas.crellen@ndm.ox.ac.uk (TC);  
ben@tropmedres.ac (BSC)

<sup>1</sup> Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, 60th Anniversary Chalermprakiat Building, 420/6 Rajvithi Road, Bangkok 10400, Thailand; <sup>2</sup> Nuffield Department of Medicine, University of Oxford, Henry Wellcome Building for Molecular Physiology, Old Road Campus, Headington, Oxford OX3 7BN; <sup>3</sup> Cambodia-Oxford Medical Research Unit, Angkor Hospital for Children, Tep Vong Road, Siem Reap, Cambodia; <sup>4</sup> Oxford University Clinical Research Unit, Centre for Tropical Medicine, 764 Vo Van Kiet, Quan 5, Ho Chi Minh City, Viet Nam

## Abstract

Multidrug-resistant *Klebsiella pneumoniae* is an increasing cause of infant mortality in developing countries. We aimed to develop a quantitative understanding of the drivers of this epidemic by estimating the effects of antibiotics on nosocomial transmission risk, comparing competing hypotheses about mechanisms of spread, and quantifying the impact of potential interventions. Using a sequence of dynamic models, we analysed data from a one-year prospective carriage study in a Cambodian neonatal intensive care unit with hyperendemic third-generation cephalosporin-resistant *K. pneumoniae*. All widely-used antibiotics except imipenem were associated with an increased daily acquisition risk, with an odds ratio for the most common combination (ampicillin + gentamicin) of 1.96 (95% CrI 1.18, 3.36). Models incorporating genomic data found that colonisation pressure was associated with a higher transmission risk, indicated sequence type heterogeneity in transmissibility, and showed that within-ward transmission was insufficient to maintain endemicity. Simulations indicated that increasing the nurse-patient ratio could be an effective intervention.

## Introduction

Infections with multidrug resistant Enterobacteriaceae constitute a major threat to public health in all regions of the world *Schwaber and Carmeli (2008); Theuretzbacher (2017)* and extended spectrum  $\beta$ -lactamase (ESBL) producing and carbapenem-resistant Enterobacteriaceae have been prioritised by the World Health Organization as pathogenic bacteria in need of novel therapeutics *World Health Organization (2014); World Health Organization (2017)*. These organisms pose the highest risk to subgroups of patients such as those undergoing surgery, requiring invasive devices, and neonates *Peleg and Hooper (2010); Goldmann (1981)*. Antimicrobial resistance is of particular

concern in developing (lower and middle-income) countries where the estimated per-capita mortality from drug-resistant bacteraemia is far greater than in high income countries and where last-line antibiotics may be unavailable or unaffordable [Lim et al. \(2016\)](#); [Zaidi et al. \(2005\)](#). The high risk of difficult-to-treat nosocomial infections threatens to undermine patient confidence in developing world hospitals and health systems [Dondorp et al. \(2018\)](#).

In this paper we focus on *Klebsiella pneumoniae*, which is amongst the most clinically important multidrug resistant Gram-negative bacteria in developing country settings [Fox-Lewis et al. \(2018\)](#); [Zellweger et al. \(2017\)](#); [Musicha et al. \(2017\)](#). Genomic studies characterising the population structure of *K. pneumoniae* have revealed a complex consisting of three separate species (*K. pneumoniae*, *K. quasipneumoniae*, *K. variicola*) that are indistinguishable by culture or standard biochemical assays [Holt et al. \(2015\)](#). Where these isolates remain undifferentiated by molecular assay, we refer to them as *K. pneumoniae sensu lato* (in the broad sense).

While there is currently widespread concern about Gram-negative bacteria as an emerging threat due to high levels of plasmid-borne resistance [Nordmann et al. \(2011\)](#), pathogens such as *K. pneumoniae* have been considered a major problem in nosocomial settings for over half a century [Yow \(1955\)](#). Research during the 1960s identified a number of drivers of *Klebsiella* colonisation and infection in hospital settings including invasive devices [Mertz et al. \(1967\)](#), environmental contamination [Kresky \(1964\)](#), introduction from the community [Kessner and Lepper \(1967\)](#), person-to-person transmission [Weil et al. \(1966\)](#), endogenous selection from antibiotic pressure [Selden et al. \(1971\)](#) and transient carriage on the hands of health-care workers [Adler et al. \(1970\)](#). In a comprehensive review of this body of work, Montgomerie concluded that ‘The likely means of transmission of *Klebsiella* is via the hands of hospital staff members’ [Montgomerie \(1979\)](#).

More recently, carriage studies in high income hospital settings in temperate regions have used whole-genome sequencing to show the critical importance of asymptomatic carriage for understanding the epidemiology of *K. pneumoniae*, establishing a firm link between gastrointestinal carriage and clinical infection [Martin et al. \(2016\)](#); [Gorrie et al. \(2017\)](#); [Gorrie et al. \(2018\)](#). In contrast to studies in high income countries where multidrug resistant *K. pneumoniae* is typically rare, prospective carriage studies in hospitalised paediatric populations in developing countries in Africa and Asia have reported hyperendemic levels of ESBL-producing Enterobacteriaceae including *K. pneumoniae* [Andriatahina et al. \(2010\)](#); [Roberts et al. \(2019\)](#); [Founou et al. \(2019\)](#); [Turner et al. \(2016\)](#) consistent with the greater burden of disease due to these organisms in lower income settings [Lim et al. \(2016\)](#); [Musicha et al. \(2017\)](#).

Despite their clinical importance, there are major gaps in our knowledge of the epidemiology of multidrug resistant *K. pneumoniae*. First, while a number of recent investigations of localised hospital outbreaks in high income settings have provided evidence that long-term environmental contamination of sinks and other sites may play a role [Clarivet et al. \(2016\)](#); [Decraene et al. \(2018\)](#); [Mathers et al. \(2018\)](#), the relative importance of persistently contaminated point sources versus patient-to-patient transmission in endemic settings remains unclear. Second, though it is widely assumed that antibiotic exposures play an important role in selecting for multidrug resistant *K. pneumoniae* [Baker et al. \(2018\)](#), such effects have not previously been quantified at the patient level in a way that disentangles antibiotic effects from general exposure to the hospital environment. Third, attempts to quantify transmissibility and how this varies by sequence type (ST) are lacking. Fourth, the impact of other factors that might plausibly affect transmission including staffing levels, infant breast feeding and use of probiotics have not been explored.

To address these knowledge gaps, we used data collected from a year-long prospective observational carriage study in a Cambodian neonatal intensive care unit and analysis methods which build upon a previously described data-driven stochastic model [Forrester and Pettitt \(2005\)](#). We fit four models with logit link functions to estimate the impact of covariates on the daily risk of acquisition of 3GC-R *K. pneumoniae s.l.*. As these models are unable to identify the force of infection, and genomic data shows the ward 3GC-R *K. pneumoniae s.l.* to be a highly heterogeneous bacterial community, we then fit five linear transmission models to estimate the force of infection for acquisition of

each ST. Finally, we use the estimated parameters to provide model-based assessments of the potential impact of hypothetical control measures. We define "acquisition" as the first detection of the organism within an infant, which may indicate transmission from other colonised infants or hospital staff, or endogenous selection as a result of antibiotic pressure.

## Results

### Descriptive Epidemiological Data

Over the year long observation period, there were consistently high rates of patient carriage of third generation cephalosporin resistant (3GC-R) *K. pneumoniae sensu lato*. Of 333 infants admitted to the neonatal unit, 121 of 289 (42%) were found to be colonised on the first swab taken within 48 hours of ward admission. A further 21 out of 44 (48%) were positive on the first swab that was taken more than 48 hours after admission. Overall, 109/191 (57%) infants who initially screened negative for 3GC-R *K. pneumoniae s.l.* became positive during their stay in the neonatal unit. Almost all 3GC-R *K. pneumoniae s.l.* isolates were ESBL producers (1412/1423; 99%), and only 5/1423 (0.35%) were resistant to imipenem. Co-colonisation with 3GC-R *E. coli* was observed in 52 infants on their first swab, and a further 102 infants became co-colonised with both resistant organisms during their stay on the neonatal unit. Full details on the study population, including blood-stream infections and mortality have been reported previously [Turner et al. \(2016\)](#), and a summary is provided in Table 1.

The daily counts of infants known to be colonised with 3GC-R *K. pneumoniae s.l.* (Figure 1A), shows no clear trend but large stochastic fluctuations which are expected given the small size of the ward (eight beds) and frequent discharges of patients and introduction of colonised infants (imported carriage). A representation of swabbing interval outcomes, as used in the models, is shown in panel B. While the median length of stay was four days, the distribution is highly skewed with a tail of long-staying patients (Figure 1C). The most frequently used antibiotic combination was ampicillin with gentamicin which was used empirically to treat suspected sepsis in infants admitted from the community and was taken on one fifth (19%) of patient days on the ward. This was followed by imipenem, which was taken on 12% of patient days. Imipenem was typically used when culture results showed non-susceptibility to first-line antibiotic choices and empirically in infants with suspected hospital-acquired infection. All other antibiotic combinations were used at a much lower frequency (Figure 1D).

### Factors Associated with Carriage Acquisition

Infants with negative cultures for 3GC-R *K. pneumoniae s.l.* on admission to the ward were prospectively screened for the organism whilst on the ward. The outcome (acquisition of the organism) was therefore interval censored between subsequent rectal swabs that were taken a median of 2 days apart (IQR 1, 3 days). In total there were 400 swab outcomes (either negatives or a first positive swab) over 864 patient days from 191 infants with a negative culture at entry.

Four models were fitted to the interval censored swab data to determine factors associated with daily risk of carriage acquisition. The best performing single intercept model (model A; Table 2) with the lowest WAIC (see Methods) considered exposure to antibiotics in the previous 96 hours and did not include a term for colonisation pressure (i.e. daily per-patient acquisition risk did not depend on the number of other patients who were colonised on a given day). Models with a 48 hour antibiotic exposure period (model B) and those that included a colonisation pressure term (model C) showed slightly worse fits.

The covariates associated with reduced daily risk of acquisition were breast milk feeding (odds ratio [OR] 0.69 [95% CrI 0.35, 1.41]) and increasing the number of nurses, for instance 3 nurses in the ward was associated with an OR of 0.55 (95% CrI 0.15, 1.77) relative to zero nurses (baseline). Male sex was also associated with a reduced risk of acquisition (OR 0.68 [95% CrI 0.43, 1.04]). Other covariates such as taking a probiotic (OR 0.88 [95% CrI 0.55, 1.41]), a severe condition (OR 1.10 [95%

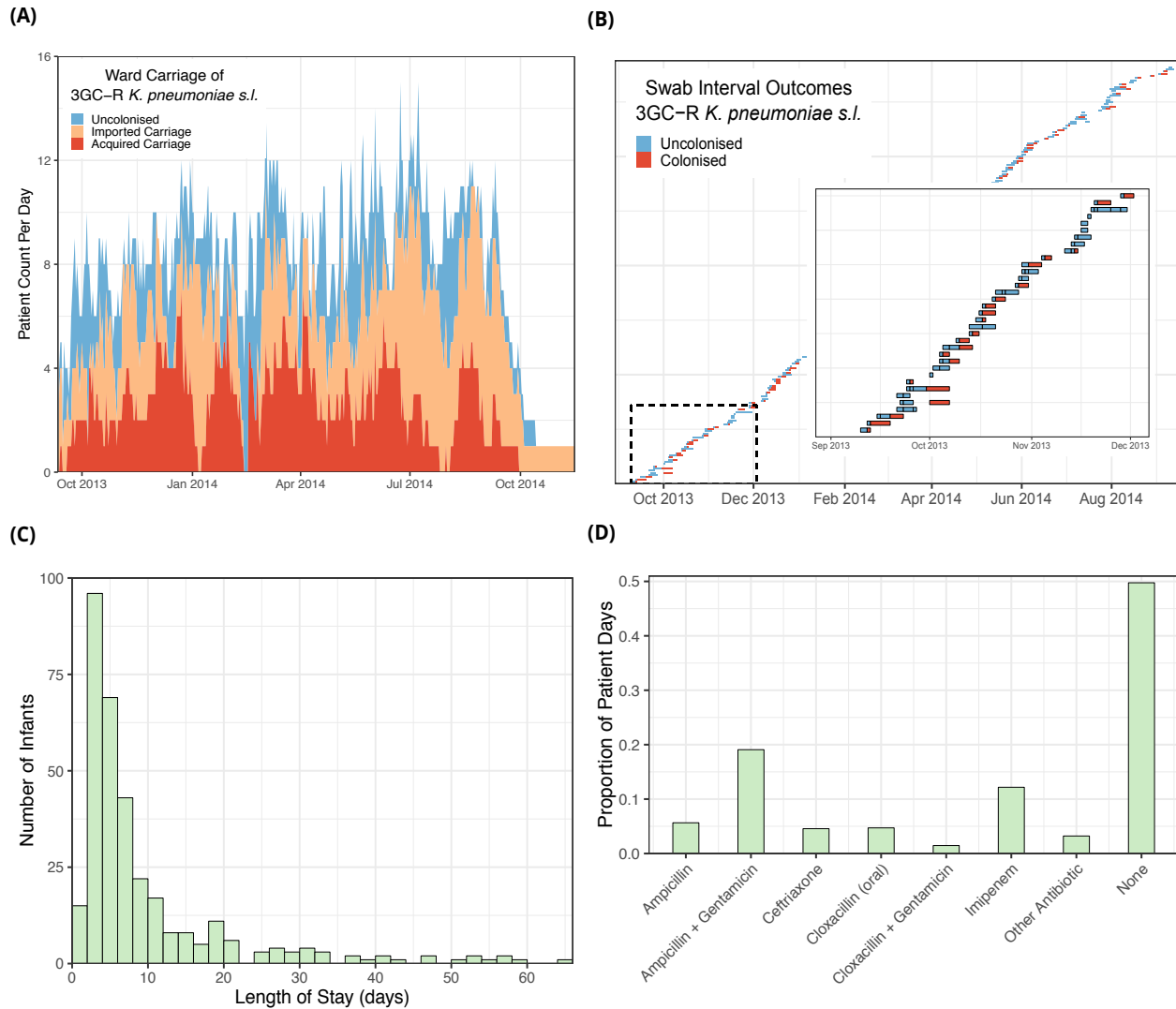
Variable	Males	Females	Total
Number of Patients	177 (53.1%)	156 (47.8%)	333 (100%)
Length of Stay in Days Median, (IQR) <sup>1</sup>	6 (4, 11)	6 (4, 11)	6 (4, 11)
Colonised with <i>K. pneumoniae</i> at Entry (or Unknown Time) <sup>2</sup>	66 (10)	55 (11)	121 (21)
Colonised with <i>K. pneumoniae</i> During Admission	54/101 (54.5%)	55/90 (61.1%)	109/191 (57.1%)
Co-colonised with <i>K. pneumoniae</i> and <i>E. coli</i> at Entry (or Unknown Time) <sup>2</sup>	26 (5)	19 (2)	45 (7)
Co-colonised with <i>K. pneumoniae</i> and <i>E. coli</i> During Admission	49/146 (33.6%)	53/135 (39.3%)	102/281 (36.3%)
Age at Entry in Days (IQR) <sup>1</sup>	8 (2, 15)	9 (1, 17)	8 (1, 16)
Probiotic Taken <sup>3</sup>	76/177 (42.9%)	62/156 (39.7%)	138/333 (41.4)
Breast Milk Fed	163/177 (92.1%)	139/156 (89.1%)	302/333 (90.1%)
Severe <sup>4</sup>	35/177 (19.8%)	32/156 (20.5%)	67/333 (20.1%)
Born Premature	30/177 (16.9%)	24/156 (15.4%)	54/333 (16.2%)

**Table 1.** Summary of characteristics of infants admitted to the neonatal intensive care unit at a children's hospital in Cambodia from September 2013 to September 2014. Colonisation status with third generation cephalosporin resistant *Klebsiella pneumoniae sensu lato* was recorded through prospectively taken rectal swabs. <sup>1</sup> Interquartile range. <sup>2</sup> Colonised at entry is defined as an initial positive swab within the first 48 hours of admission; if the first swab is positive and it was taken later than 48 hours from admission then the infant is considered to be colonised at an unknown time. <sup>3</sup> Assigned by clinician to receive oral *Lactobacillus acidophilus*. <sup>4</sup> Severe symptoms are requiring ventilation, continuous airway pressure or inotropes.

CrI 0.55, 2.13]), prior colonisation with 3GC-R *E. coli* (OR 1.07, [95% CrI 0.65, 1.75]), and age at admission (10 days compared with zero days, OR 0.95 [95% CI 0.73, 1.24]) had ORs distributions centred closer to unity. As anticipated (though, to our knowledge, not previously shown), antibiotics taken (intravenously, with the exception of cloxacillin) within the past 96 hours were mostly associated with an increased risk of colonisation with 3GC-R *K. pneumoniae s.l.*. Ampicillin (OR 1.77 [95% CrI 0.88, 3.26]), ampicillin with gentamicin (OR 1.96 [95% CrI 1.18, 3.36]), ceftriaxone (OR 1.85 [95% CrI 0.68, 4.54]), oral cloxacillin (OR 1.49 [95% CrI 0.47, 4.02]), and cloxacillin with gentamicin (OR 1.94 [95% CrI 0.55, 5.66]) were all associated with an increased risk of acquisition. Only intravenous imipenem (OR 1.01 [0.40, 2.30]) had a posterior distribution centred near unity, consistent with the carbapenem sensitivity of 3GC-R *K. pneumoniae* found in this setting. See Figure 2A for odds ratio posterior distributions.

Using the covariate posterior distributions, we also estimated the probability of colonisation for each of the 864 patient days where patients were at risk for acquiring 3GC-R *K. pneumoniae*





**Figure 1.** Descriptive epidemiological data from a cohort of 333 infants admitted to a neonatal unit in a Children’s Hospital in Cambodia from September 2013 to September 2014. Daily counts of neonates colonised with third generation cephalosporin resistant (3GC-R) *Klebsiella pneumoniae sensu lato* over the study period are shown in panel A, where colour reflects uncolonised, imported or acquired cases, according to case definitions. The total height of the peaks shows the ward occupancy on that day. The results from rectal swabs among the 191 infants uncolonised at entry for 3GC-R *K. pneumoniae* s.l. are shown in panel B, with the window highlighting the swab outcomes from the first thirty five infants uncolonised at entry. Each row represents a patient and each coloured block represents a swab interval, where the width is the number of days in the interval (i.e. time between swabs). Outcomes are shown up to the first swab positive for 3GC-R *K. pneumoniae* s.l., after which time the patient is assumed to be colonised until discharge. The length of stay distribution for infants in the neonatal unit is shown as a histogram in panel C, where the bin width is two days. An infant’s length of stay is the total time in the neonatal unit during the study period, including re-admissions. The 333 infants were present in the neonatal unit for a total of 3417 study days. The proportion of study days when infants took the six most common antibiotic combinations, or other antibiotics, or none are shown in panel D.

s.l.. The median daily probability of first acquisition of 3GC-R *K. pneumoniae* s.l. for an infant was estimated from the best fitting model as 0.15. There is considerable variability in the risk of acquisition between patient days and the medians of the posterior probability distribution ranges nearly eight-fold from 0.047 to 0.35 (Figure 2B).

The risk of becoming colonised with 3GC-R *K. pneumoniae* s.l. is cumulative over an infant’s length of stay and varies in response to interventions, such as consumption of antibiotics. We

show the cumulative risk of first acquisition under two scenarios: 1) where an initially four day old, breastfed, female infant remains in the ward for eight days without taking antibiotics or probiotics; and 2) where an infant with the same characteristics is prescribed ampicillin with gentamicin from day three onwards. The median cumulative risk of acquiring 3GC-R *K. pneumoniae s.l.* after eight days for the first scenario is 0.57 (80% CrI 0.42, 0.71) and for the second scenario is 0.75 (80% CrI 0.64, 0.85). Although the median cumulative risk between the two scenarios diverges the longer the infants are in the neonatal unit, the uncertainty also increases with time (Figure 2C).

### Swab Sensitivity

We estimated the sensitivity of rectal swabs for detecting 3GC-R *K. pneumoniae s.l.* by examining the swabs that followed a positive from the same patient. There were 936 such swabs which were taken from a patient after at least one swab positive for 3GC-R *K. pneumoniae s.l.*, and 90 (9.6%) of these were negatives. Under the assumption that all negatives following a positive culture are false negatives, the false negative rate posterior median was 0.096 (95% CrI 0.078, 0.12) and the posterior median swab sensitivity was 0.90 (95% CrI 0.88, 0.92). Under the second assumption that three or more consecutive negative swabs following a positive culture represent a true decolonisation event, there were 72 false negatives, giving a false negative rate of 0.073 (95% CrI 0.058, 0.091) and a swab sensitivity of 0.93 (95% CrI 0.91, 0.94).

### Model Assessment and Comparison

The measures of Markov chain convergence showed high effective sample sizes (>400) and  $\hat{R} < 1.01$ , indicating that the chains had run for long enough and had mixed well (see Methods and Figure 2- figure supplement 1). Model assessment was performed with a posterior predictive check; we estimated the probability of acquisition for each of the 400 swabbing intervals and binning these probabilities into groups defined by the quintiles. We then calculated the expected number of colonisation events in each of the five groups and compared these with the observed number of acquisitions. Within each of these groups, the posterior median of the predicted number of acquisitions was close to the observed number of events, and the observed values were always within the 80% CrI of the model estimates (Figure 2D). The results from fitting risk factor model 1 with alternative prior distributions are shown in Figure 2- figure supplement 2; substantially reducing the variance of the priors for the model intercept and covariates has a negligible effect on the posterior parameter estimates.

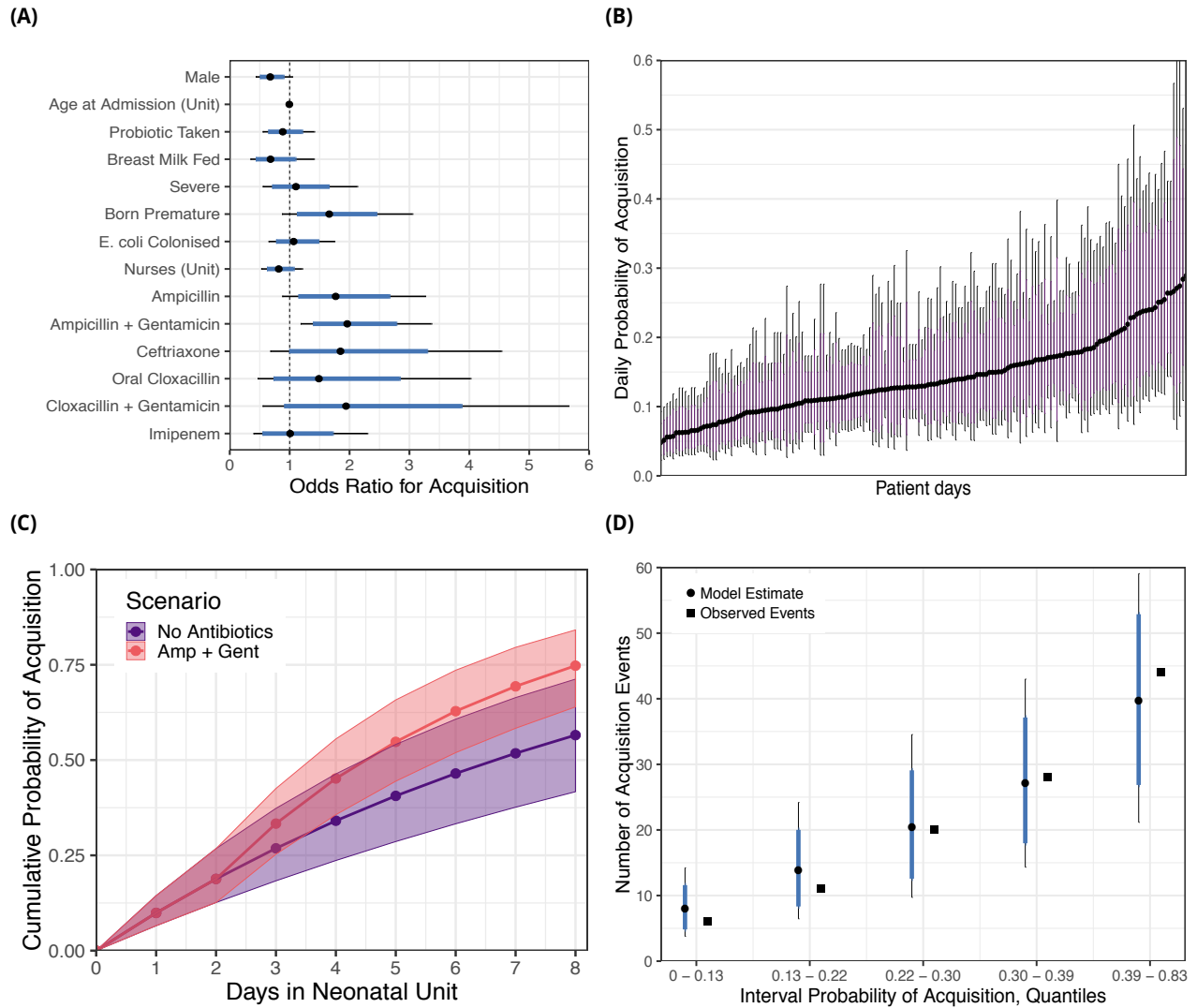
When the intercept was permitted to vary by study month in a hierarchical model (risk factor model D; Table 2), little variation was observed between months; the median posterior baseline probability ranged from 0.20 to 0.23 with wide credible intervals. As these models did not include a colonisation pressure term, the intercept incorporated time-varying changes in the underlying intensity of transmission. The low variance in the monthly intercepts therefore suggests a relatively constant force of infection over the 12 month study period. In models where we included a colonisation pressure term (risk factor model C; Table 2) this was found to have a slightly negative slope for acquisition of 3GC-R *K. pneumoniae s.l.* (OR 0.96 [95% CrI 0.86, 1.09]). This is surprising, as if patient-to-patient transmission was occurring, we would usually expect the force of infection to increase with the colonisation pressure [Bonten \(2012\)](#). The finding therefore suggested that one of the following three possibilities was true: i) patient-to-patient transmission was not occurring at a high frequency in this ward; ii) patient-to-patient transmission was occurring but, because of the continually high ward-level prevalence, variations in the force of infection could not be identified; iii) patient-to-patient transmission was occurring but exposure to the presence of two or more colonised patients presented a similar risk for acquisition as exposure to one. We therefore used *K. pneumoniae s.l.* whole-genome sequence data to help determine the most plausible scenario.

Risk Factor Model	Parameters	Priors	Posterior Median (95% CrI) <sup>1</sup>	WAIC <sup>2</sup>
(A) Single intercept Standard covariates <sup>3</sup> 96 hour antibiotic exposure	$\alpha$ (intercept) $\beta$ (slopes)	normal(0, 10) normal(0, 5)	0.23 (0.055, 0.60) ORs <sup>4</sup> in results	438
(B) Single intercept Standard covariates <sup>3</sup> 48 hour antibiotic exposure	$\alpha$ (intercept) $\beta$ (slopes)	normal(0, 10) normal(0, 5)	0.26 (0.068, 0.63) Not shown	441
(C) Single intercept Standard covariates <sup>3</sup> + colonisation pressure term <sup>5</sup> 96 hour antibiotic exposure	$\alpha$ (intercept) $\beta$ (slopes)	normal(0, 10) normal(0, 5)	0.26 (0.059, 0.64) Not shown	440
(D) Intercept varies by month Standard covariates <sup>3</sup> 96 hour antibiotic exposure	$\alpha$ [month] (intercept) $\mu$ (normal mean) $\sigma$ (normal standard deviation) $\beta$ (slopes)	normal( $\mu$ , $\sigma$ ) normal(0, 3) half-normal(0, 1) normal(0, 3)	Varies by month <sup>6</sup> 0.21 (0.044, 0.57) 0.54 (0.51, 0.63) Not shown	440

**Table 2.** Comparison of models for the risk of acquiring third generation cephalosporin resistant *Klebsiella pneumoniae sensu lato* over 864 patient days in a neonatal intensive care unit in Cambodia. Models vary by explanatory variables (A-C) or by permitting the intercept to vary between study months in a hierarchical model (D). Models were fitted on the log-odds scale with a logit link function, hence prior distributions are shown as log-odds. Posterior parameter distributions have been transformed using the logistic function and are shown as probabilities. Prior distributions are normal distributions, shown in brackets are the the mean and standard deviation respectively. <sup>1</sup> 95% Credible interval. <sup>2</sup> Widely applicable information criterion (a model comparison statistic where lower values indicate better fitting models). <sup>3</sup> Standard covariates: use of ampicillin, ampicillin + gentamicin, cloxacillin (oral), ceftriaxone, cloxacillin + gentamicin, and imipenem within the previous 48 or 96 hours; whether breast fed; receipt of an oral probiotic on entry (*Lactobacillus acidophilus*), sex, premature (born before the 37th week of pregnancy), severity (defined as severe if requiring ventilation, continuous positive airway pressure or intubation), already colonised with 3GC-R *E. coli*, age in days on first admission to the NU, and the daily number of nurses on the ward. These explanatory variables were treated as binary and, where appropriate, time-varying. Covariates were recorded for every day the infant was present in the neonatal unit (see Methods for full details). <sup>4</sup> Odds ratios. <sup>5</sup> Colonisation pressure is the number of known colonised patients on the ward on a given day. <sup>6</sup> Median posterior probability ranges by month 0.20-0.23.

### *Klebsiella* Whole-Genome Assemblies

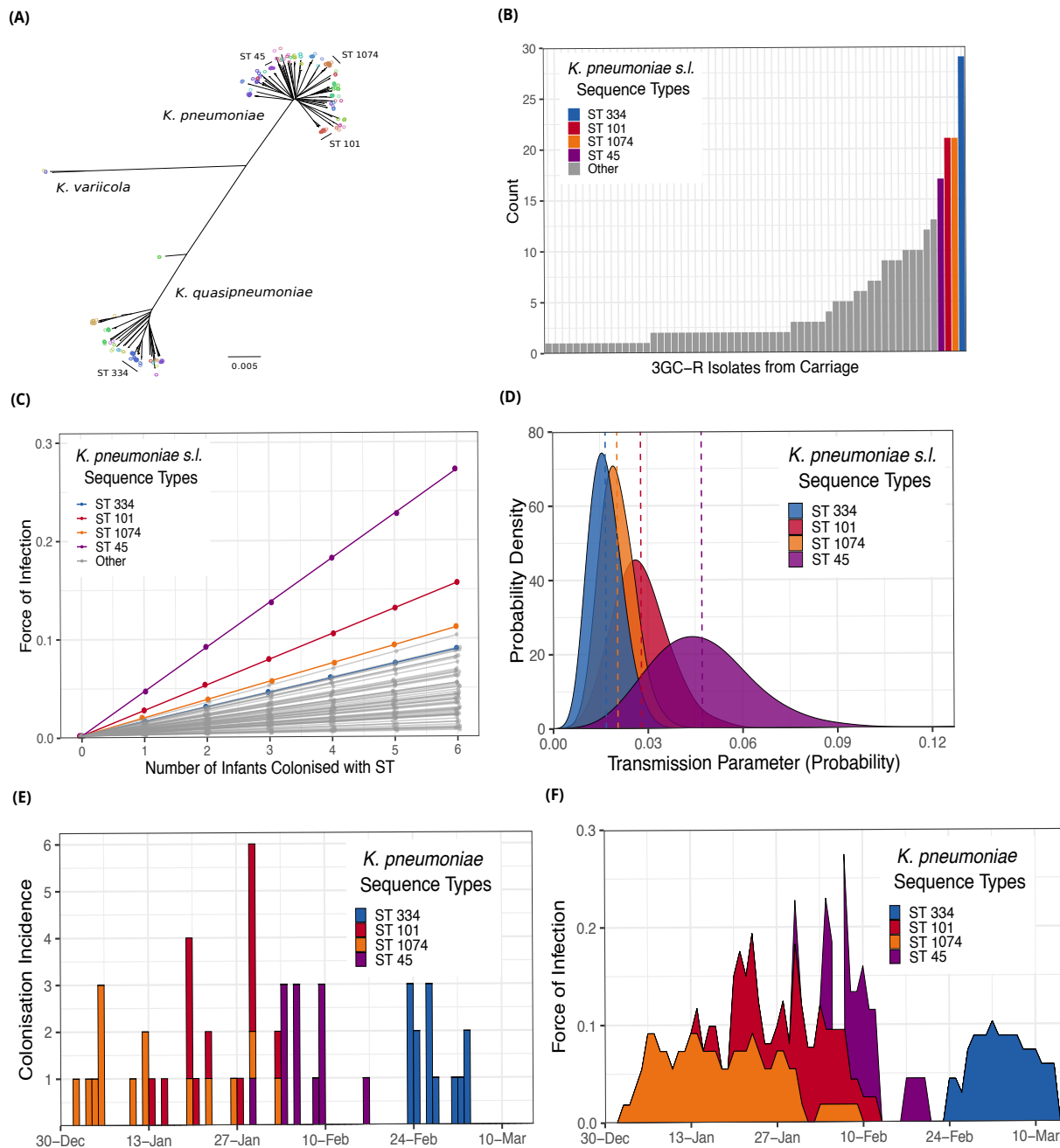
We examined whole-genome assemblies of 317 3GC-R *K. pneumoniae s.l.* isolates cultured from rectal or environmental swabs in the neonatal unit over a four month period (see Methods). A phylogeny based on k-mer distances between assemblies is shown in Figure 3A. Of note is the highly diverse and structured nature of the pathogen population, in contrast to one dominated by a clonal expansion of a single lineage. Overall 62 distinct sequence types were identified in our collection of isolates. The species identified from culture as *K. pneumoniae s.l.* consists of three distinct subpopulations that meet the criteria for separate species. We isolated all three species from infants in the cohort (*K. pneumoniae*  $n=219$ , *K. quasipneumoniae* subspecies *similipneumoniae*  $n=95$ , *K. variicola*  $n=3$ ), and found diversity similar to that observed in a global collection of *K. pneumoniae* isolates [Holt et al. \(2015\)](#), suggesting that the diversity accumulated within a Cambodian neonatal unit over four months is comparable to the diversity of *K. pneumoniae* globally. Many STs were characterised by only a single carriage isolate, suggestive of importations that were not subsequently transmitted to other patients (Figure 3B). The STs with the largest number of carriage isolates were ST334 (*K. quasipneumoniae*;  $n=29$ ), ST101 (*K. pneumoniae*;  $n=21$ ), ST1074 (*K. pneumoniae*;  $n=21$ ) and ST45 (*K. pneumoniae*;  $n=17$ ). The most frequent *bla*<sub>ESBL</sub> genes in the whole-genome assemblies were CTX-M-15 (201/317; 63%), followed by CTX-M-14 (43/317; 13%) and CTX-M-63 (38/317; 12%).



**Figure 2.** Posterior distributions for risk factors for the daily probability of acquiring third-generation cephalosporin resistant (3GC-R) *Klebsiella pneumoniae sensu lato* among 191 susceptible neonates. Odds ratios for the daily risk of colonisation are shown in panel A. The daily risk of colonisation per patient day is shown in panel B. Note that the 864 patient days have been thinned by a factor of five for visualisation. The cumulative risk for different patient scenarios is explored in panel C; a four day old girl, born full term, without severe conditions, breast milk fed and not taking antibiotics or probiotics over eight days in the neonatal unit is shown in blue. The red line shows the same infant, however ampicillin + gentamicin is taken from day three onwards. The lines and points in both cases show the cumulative probability posterior median, and the shaded area shows the 80% credible interval (CrI). In panel D, we took the probability of colonisation for each of the 400 swab interval and binned them into five quantiles. We then compared the expected number of colonisation events predicted by the model with the observed number of colonisation events (squares) in the swab intervals by quantile. In panels A, B and D points represent posterior medians, thick blue / purple lines represents the 80% CrI and thinner black lines represent the 95% CrI.

## Transmission Models for Sequence Types

We fitted mechanistic models representing different transmission processes to the ST swab data. Within the four month period where sequence data were available, there were 171 events for first acquisition of a 3GC-R *K. pneumoniae s.l.* ST among 150 infants. Between transmission models 1-3 that were initially tested, the model with the best fit to data by WAIC was transmission model 2 (see Table 3 and Methods), which has an intercept ( $\alpha$ ), representing a constant risk of acquisition, and a slope ( $\beta$ ) which scales the risk of acquisition for each infant colonised with a given ST in the ward



**Figure 3.** Population structure of third-generation cephalosporin-resistant (3GC-R) *Klebsiella* and force of infection by sequence type (ST). An unrooted phylogeny of 317 3GC-R *Klebsiella* isolates cultured from rectal and environmental swabs over a four month period in a neonatal unit in a children's hospital in Cambodia is shown in panel A, where the branch lengths correspond to the mash distance (a measure of k-mer similarity) between whole-genome assemblies. The four largest STs are labelled as well as the population subdivisions by *Klebsiella* species. The frequency distribution of STs is shown in panel B, with the four largest STs shown in colour. Results from a transmission model estimating the force of infection by ST are shown in panel C, where the force of infection scales linearly with the number of colonised infants with that ST. The largest four STs have again been highlighted. Horizontal jitter has been applied to prevent overplotting of points. The uncertainty around the transmission parameter estimates are shown in Panel D for the four most common STs, where the posterior mean is shown with a dotted line. The daily incidence of new colonisation events with the four most frequent STs are shown between the 1st January to the 15th March 2014 in panel E, along with the estimated force of infection over the same period in panel F using parameter estimates of  $\beta$  from transmission model 4 (Table 3).

(i.e. it accounts for colonisation pressure). The model estimated the values for  $\alpha$  as 0.0019 (95% CrI 0.0015, 0.0023) and  $\beta$  as 0.0097 (95% CrI 0.0075, 0.012). We then fitted transmission model 2 with a random effect term, where  $\beta$  was permitted to vary by ST and the underlying distribution of  $\beta$  was assumed to follow a beta distribution, with shape hyper-parameters  $\alpha_s$  and  $\beta_s$  (transmission model 4). We also fitted transmission model 2 with a random effect term where  $\alpha$  was permitted to vary by ST (transmission model 5).

Transmission model 1 has a constant colonisation pressure by ST which was unlinked to the daily number of colonised individuals with that ST; this model showed a substantially worse fit to the data by WAIC (Table 3). Transmission model 3, which included terms for colonisation pressure and contamination in the hospital environment also failed to improve the model fit, and the high estimate for  $\lambda$  (4.7 95% CrI [2.0, 7.0]) suggests that contamination left by previously colonised infants decays rapidly to background levels, with an estimated environmental half life of 3.6 hours (95% CrI 2.4, 7.6 hours; Table 4). Varying the  $\alpha$  parameter by ST (transmission model 5) resulted in a substantially worse fit to data, suggesting there was not enough information in the model to differentiate ST-specific background rates of colonisation.

The central estimates of the force of infection by ST from transmission model 4 are shown in Figure 3C, with estimates from the four most frequent STs highlighted in colour. The uncertainty around these parameter estimates for the four STs is shown in Figure 3D. The daily interval-censored colonisation incidence for the most frequent STs are clustered in time, generally emerging and reaching extinction in the ward within a matter of weeks, suggestive of importation and subsequent patient-to-patient transmission. The incidence and estimated force of infection for the four most frequent STs over a period in the study where all 3GC-R *K. pneumoniae s.l.* isolates were sequenced are shown in Figures 3E and 3F. Parameter estimates for all transmission models are shown in Table 3. Model fitting diagnostics showed that Markov chains had converged satisfactorily (see Methods and Figure 3- figure supplement 1).

### Sequence Type SNP Diversity

We mapped reads from isolates in two STs with the highest estimated force of infection (ST45 and ST101; see Figure 3) to ST consensus reference genomes. The 21 ST101 carriage isolates had a mean read depth ranging from 25x to 125x (median 57x), and the 17 ST45 carriage isolates had a mean read depth ranging from 20x to 67x (median 56x). All the carriage isolates in both STs had >90% of the genome covered by >5x coverage. We then called and filtered SNPs (see Methods) to determine if the relatedness of carriage isolates within STs was consistent with recent person-to-person transmission. The pairwise number of variants in ST101 isolates between infants ranged from 15 to 68 SNPs (median 30), which was comparable to the variation seen within infants in ST101 (from 18 to 38 SNPs; median 28). Similarly in ST45 the pairwise SNP differences between infants varied from 13 to 223 (median 125), which was comparable to the within-host ST diversity (from 55 to 212 SNPs; median 124). Therefore the SNP diversity observed within and between-hosts was very similar for both ST45 and ST101.

### Simulations with Agent-Based Models

We used the posterior parameter estimates obtained from model fitting for forward simulations using a dynamic agent-based model in order to evaluate the potential impact of interventions (see Methods). We first estimated the ward-level reproduction number ( $R_A$ ) for 3GC-R *K. pneumoniae s.l.* by simulating the introduction of a single colonised patient into a ward of eight susceptible patients. For all patients, length of stay was sampled from the empirical length of stay distribution (Figure 1C) and colonised patients had a transmission potential sampled from the posterior hyperparameter distribution from transmission model 4 (Table 3). The median of the  $R_A$  distribution was 0.65 and 95% of values fell between 0.36 and 1.09. The distribution of  $R_A$  values is shown in Figure 4A.

We then simulated the impact of interventions to reduce the rate of 3GC-R *K. pneumoniae s.l.* acquisition by combining parameter estimates for colonisation pressure from transmission model

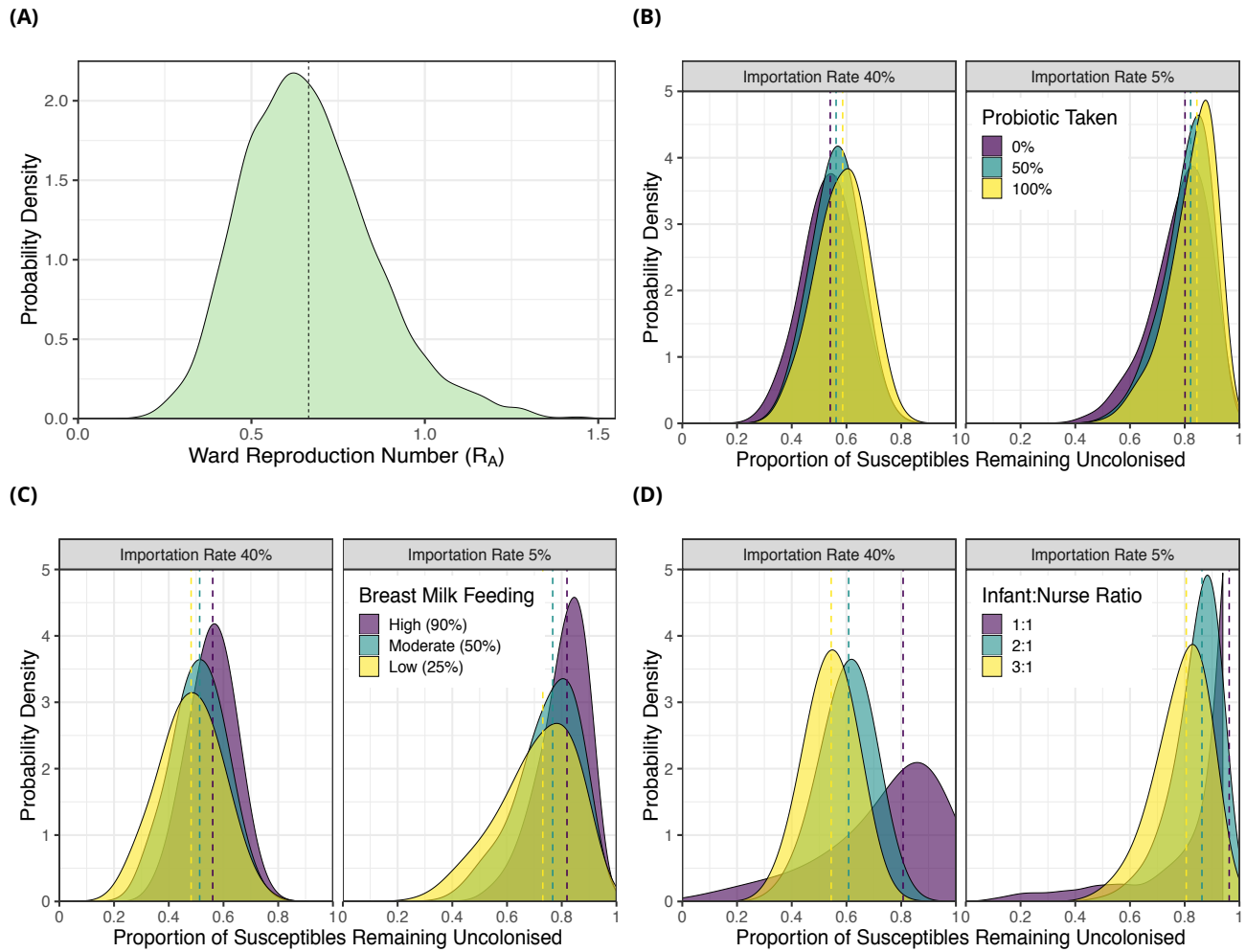


Transmission Model (Equations in Methods)	Parameters	Priors	Posterior Median (95% CrI) <sup>1</sup>	WAIC <sup>2</sup>
(1) Constant risk of transmission	$\alpha$ (intercept)	beta(2, 8)	0.0038 (0.0032, 0.0044)	1919
(2) Pseudo mass action (PMA) principal	$\alpha$ (intercept) $\beta$ (transmission) <sup>3</sup>	beta(2, 8) beta(2, 8)	0.0019 (0.0015, 0.0024) 0.0096 (0.0075, 0.012)	1739
(3) PMA plus environmental contamination, from colonised patients, which decays over time	$\alpha$ (intercept) $\beta$ (transmission) <sup>3</sup> $\gamma$ (environment) $\lambda$ (decay term)	beta(2, 8) beta(2, 8) beta(2, 8) half-normal(1, 2)	0.0019 (0.0014, 0.0023) 0.0088 (0.0063, 0.011) 0.097 (0.0086, 0.37) 4.7 (2.2, 7.1)	1740
(4) Hierarchical PMA varying transmission coefficient by ST <sup>4</sup>	$\alpha$ (intercept) $\beta$ (transmission) <sup>3</sup> $\alpha_s$ (beta shape 1) $\beta_s$ (beta shape 2)	beta(2, 8) beta( $\alpha_s$ , $\beta_s$ ) half-normal(2, 5) half-normal(8, 5)	0.0021 (0.0016, 0.0026) Varies by ST <sup>4</sup> 0.23 (0.14, 0.38) 17 (9.7, 25)	1733
(5) Hierarchical PMA varying intercept by ST <sup>4</sup>	$\alpha$ (intercept) $\beta$ (transmission) <sup>3</sup> $\alpha_s$ (beta shape 1) $\beta_s$ (beta shape 2)	beta( $\alpha_s$ , $\beta_s$ ) beta(2, 8) half-normal(2, 5) half-normal(8, 5)	Varies by ST <sup>4</sup> 0.0094 (0.0073, 0.012) 0.23 (0.16, 0.31) 21 (13, 29)	1793

**Table 3.** Transmission models fitted to prospectively collected, genotyped swab data on the acquisition of third-generation cephalosporin-resistant (3GC-R) *Klebsiella pneumoniae sensu lato*. The table shows the parameters, prior and posterior distributions along with the WAIC (model comparison measure where lower values indicate a better fit to data). See methods for equations. Normal prior distributions show the mean and standard deviation respectively within brackets, beta prior distributions show the two shape parameters within brackets. <sup>1</sup> 95% Credible interval. <sup>2</sup> Widely applicable information criterion (WAIC). <sup>3</sup> Transmission parameter that is multiplied by the number of infants colonised with the sequence type on the same day to give the force of infection (colonisation pressure). <sup>4</sup> 3GC-R *K. pneumoniae s.l.* sequence type (ST).

4 with the marginal effect of modifiable covariates from risk factor model A (Table 2). In the first intervention scenario we varied the proportion of infants given an oral probiotic, in addition to varying the proportion of infants that were colonised on entry (imported cases; 5% or 40%). We used as an outcome the proportion of infants susceptible to 3GC-R *K. pneumoniae s.l.* on admission that remained uncolonised on discharge. When the importation rate was high (40%; similar to the rate in our study population, see Table 1), setting the proportion of infants taking the probiotic to be 0%, 50% or 100% resulted in the median proportion remaining uncolonised as 0.54 (95% interval 0.34, 0.72), 0.56 (95% interval 0.38, 0.72) and 0.59 (95% interval 0.39, 0.75) respectively. In the lower importation setting (5% of infants colonised on entry), setting the proportion of infants taking the probiotic to be 0%, 50% or 100% resulted in the median proportion remaining uncolonised as 0.80 (95% interval 0.54, 0.93), 0.82 (95% interval 0.60, 0.93) and 0.84 (95% interval 0.61, 0.94) respectively.

In the second intervention scenario, we varied the proportion of breast milk fed infants between 25%, 50% and 90%, in addition to varying the proportion of infants that were categorised as imported cases. When the importation rate was high (40%), altering the proportion of infants breast fed between 25%, 50% and 90% resulted in the median proportion remaining uncolonised as 0.48 (95% interval 0.25, 0.69), 0.51 (95% interval 0.31, 0.70), and 0.56 (95% interval 0.38, 0.72) respectively. When the importation rate is low (5%) altering the proportion of infants breast fed between 25%, 50% and 90% resulted in the median proportion remaining uncolonised as 0.73 (95% interval 0.40, 0.92), 0.77 (95% interval 0.48, 0.92) and 0.84 (95% interval 0.61, 0.94) respectively.



**Figure 4.** Simulation results from dynamic agent based models using parameter estimates on acquisition of third generation cephalosporin resistant *Klebsiella pneumoniae sensu lato* among neonates in a Children's Hospital in Cambodia. The distribution of  $R_A$  values was obtained by taking 2000 samples from the force of infection posterior distribution, and for each sample running the agent-based simulation 100 times and taking the mean value. The results from simulating counterfactual scenarios with a dynamic agent based model are shown in panels B and C. In B, the proportion of infants breast milk fed was varied between 0.25, 0.5 and 0.9 in settings with a high proportion of imported cases (0.4) and a lower proportion of imported cases (0.05). In C, the infant nurse ratio was varied between 3:1, 2:1 and 1:1 in settings with a high proportion of imported cases (0.4) and a lower proportion of imported cases (0.05). The outcome measure in all simulations is the proportion of infants susceptible on entry that remained uncolonised on discharge. The simulated outcomes are displayed as density plots, with dashed lines showing the median.

296 In the third scenario we simulated either three, four or eight nurses in the ward each day,  
 297 corresponding to infant:nurse ratios of roughly 3:1, 2:1 and 1:1 respectively. Again, we examined  
 298 this effect in settings with different importation rates. Varying the number of nurses in the ward  
 299 between three, four and eight in the high importation setting resulted in median proportions of  
 300 initially uncolonised infants who remained uncolonised throughout their neonatal unit stay of 0.54  
 301 (95% interval 0.36, 0.71), 0.61 (95% interval 0.39, 0.78) and 0.81 (95% interval 0.21, 0.97) respectively.  
 302 In the lower importation setting, varying the infant:nurse ratio resulted in the median proportions  
 303 remaining uncolonised of 0.81 (95% interval 0.56, 0.93), 0.86 (95% interval 0.62, 0.95) and 0.96 (95%  
 304 interval 0.32, 0.99) respectively. Of all the simulated interventions therefore, increasing the number  
 305 of nurses on the ward had largest impact on reducing colonisation rates. The distributions of the  
 306 outcome variables from all simulations are shown in Figure 4B, 4C and 4D.

Parameter	Method	Estimate	Uncertainty interval	Key Assumptions
Daily risk of acquisition for neonates	Bayesian Regression Model	0.15	0.091, 0.19 (IQR <sup>1</sup> )	Culture diagnostic 100% sensitive
Force of infection from one colonised infant	Bayesian Transmission Model	0.016	0.0093, 0.027 (95% CrI <sup>2</sup> )	Expected values from transmission model 4
Swab sensitivity (1)	Negatives following a positive swab Beta conjugate prior	0.90	0.88, 0.92 (95% CrI <sup>2</sup> )	All positives are false negatives Beta(1,7) prior
Swab sensitivity (2)	Negatives following a positive swab Beta conjugate prior	0.93	0.91, 0.94 (95% CrI <sup>2</sup> )	Three consecutive negatives are a true decolonisation Beta(1,7) prior
Environmental half life <sup>3</sup>	Bayesian Transmission model	3.6 hours	2.4, 7.6 hours (95% CrI <sup>1</sup> )	Exponential decay Normal(1, 2) prior
Ward reproduction number $R_A$ )	Agent based simulation	0.65	0.36, 1.1 (95% interval <sup>4</sup> )	Ward size of 8 susceptible neonates

**Table 4.** Key epidemiological parameters estimated in this study from longitudinal swab data on third generation cephalosporin resistant *Klebsiella pneumoniae sensu lato* from a neonatal intensive care unit from a Children's Hospital in Cambodia. <sup>1</sup> Interquartile range (IQR) taken from distribution of daily risk of acquisition (Figure 2B). <sup>2</sup> Credible interval (CrI). <sup>3</sup> Inverse rate of decay of environmental contamination, as estimated in transmission model 3. <sup>4</sup> 95% of simulated values fell within this interval.

## Discussion

In a hyperendemic, developing country hospital setting, we analysed the transmission dynamics of 3GC-R *K. pneumoniae s.l.* prospectively over one year. We compared the support for competing hypotheses about modes of spread, quantified effects of antibiotic exposures as drivers of the epidemic, evaluated risk factors for transmission, and forward simulated to evaluate the potential benefits of interventions.

We found that carriage strains of 3GC-R *K. pneumoniae s.l.* among neonates constituted a highly diverse population, with considerable intra-host variation within STs. There were frequent acquisitions or detection of resistant *K. pneumoniae s.l.* that were closely related to strains carried by other infants on the ward at the time of acquisition (Figure 3E). Within-host diversity of these lineages was similar to between-host diversity in potentially linked cases. Moreover, once genomic information was considered, models incorporating colonisation pressure as a risk factor for acquisition showed substantially better fits to data than models without colonisation pressure (Table 3). Taken together with the lack of persistent environmental contamination *Smit et al. (2018)* and the lack of improvement in model fit when long-term environmental contamination was considered, these findings add support to the view that patient-to-patient transmission (much of which is likely to be mediated by contacts with healthcare workers) is the main driver of resistant *K. pneumoniae* acquisition, and that colonised patients represent the primary reservoir. Notably, this result was only apparent when we incorporated genomic data into our models to estimate transmission parameters by ST.

We are aware of one other prospective study of ESBL-producing Enterobacteriaceae colonisation dynamics from a developing country **Bonneault et al. (2019)**. This study considered ESBL-producing Enterobacteriaceae colonisation in a neonatal unit in Madagascar over a period of six months, and found similar rates of acquisition with ESBL-producing Enterobacteriaceae to those seen here. Though sequencing data were not used, fitting transmission models to data provided evidence of patient-to-patient and healthcare worker-patient transmission, particularly for ESBL-producing *K. pneumoniae* and the estimated daily transmission parameter (0.05; 0.008, 0.14) is similar to our estimates for the most common STs (Figure 3D).

Two other prospective carriage studies of *K. pneumoniae* have used whole genome sequencing to identify possible transmission events: while a one year study in an adult intensive care unit in Australia found five epidemiologically plausible intra-hospital transmission chains **Gorrie et al. (2017)**, a one year carriage study in two geriatric wards (also in Australia) found no evidence of patient-to-patient transmission chains **Gorrie et al. (2018)**. Other studies using genomic epidemiology to study resistant *K. pneumoniae* transmission have been performed in adult wards in high income settings **Snitkin et al. (2012)**; **Haller et al. (2015)**; **Snitkin et al. (2017)**. While such studies have also provided support for patient-to-patient transmission playing an important role, such retrospective investigations made in response to reported outbreaks of multidrug resistant *K. pneumoniae* might not reflect typical patterns of transmission. A study using proximity sensors to investigate transmission of ESBL-producing Enterobacteriaceae over four months in a long term care facility in France, found stronger support for person-to-person transmission as the main route of acquisition for *K. pneumoniae*, though the evidence for person-to-person transmission of ESBL-producing *E. coli* was weaker **Duval et al. (2019)**. This suggests that separate mechanisms may drive the transmission of different ESBL-producing organisms and that this should be considered when analysing carriage data from multiple species of Enterobacteriaceae.

Under the assumption that patient-to-patient transmission was driving the epidemic, we calculated that the rates of transmission within the ward were insufficient to maintain endemic transmission (i.e. the ward-level reproduction number,  $R_A < 1$ ) **Anderson and May (1992)**. These findings are comparable to a study which determined the relative force of infection between ESBL-producing *E. coli* and other ESBL-producing Enterobacteriaceae in 13 European intensive care units, finding that the latter (mainly *K. pneumoniae*) had a transmission capacity almost 3 times greater than the former, and that the single-admission reproduction number for both classes of organisms were well below one **Gurieva et al. (2017)**. Our central estimate of  $R_A$  is, however, substantially higher (0.65 compared with 0.17); this difference in estimated transmission potential may reflect differences in staff-to-patient ratios, different standards of hygiene and infection control, different patterns of antibiotic use, and differences in the patient population **Dondorp et al. (2018)**. The findings from both studies indicate that repeated importation into the unit is needed to maintain endemicity of resistant *Klebsiella*. Imported cases may be acquired from other wards within the same hospital, other hospitals within the referral network, or community transmission. When rates of importation are high, as observed in this setting where up to 43% of infants may have been colonised on initial ward admission (Table 1), even effective interventions are limited in how many acquisition events they can prevent due to a high underlying colonisation pressure (Figure 4). The public health implications are therefore that control measures should be coordinated regionally and targeted to the wider hospital patient referral network **Donker et al. (2012)**; **Ciccolini et al. (2013)**.

Amongst the most important findings was the consistent association between a patient's antibiotic exposure and an increased risk of acquiring 3GC-R *K. pneumoniae* s.l. or detection of the organism due to within-host selection. To our knowledge, the role of antibiotics as drivers of carriage dynamics has not previously been explored with appropriate methods to account for time-dependent antibiotic exposures in Gram-negative bacteria, but there are reasons for believing it is likely to be a key mechanism through which antibiotics select for ESBL *K. pneumoniae* **Tedijanto et al. (2018)**. With the exception of imipenem, for which there was no association with the risk of acquisition for these predominantly carbapenem-susceptible bacteria (1418/1423 *K.*

378 *pneumoniae* s.l. isolates sensitive to imipenem; 99.6%), effect sizes were similar for different  
 379 antibiotic combinations (median posterior OR ~2; Figure 2A). The narrower credible interval for the  
 380 ampicillin + gentamicin combination (OR 1.96, 95% CrI 1.18, 3.36) reflects the much higher usage  
 381 of this antibiotic combination compared to others (Figure 1D). These effects are consistent with  
 382 hypotheses about antibiotic therapy leading to reduced microbiome diversity and subsequently  
 383 increasing the risk of colonisation with drug resistant bacteria, which face less competition from  
 384 fitter, sensitive strains *Lipsitch et al. (2000)*. Microbiome analysis has shown that the greatest  
 385 dysbiosis following antibiotic therapy in healthy adults is four days after treatment starts *Palleja*  
 386 *et al. (2018)*, lending confidence to our model comparison selecting a 96 hour exposure period over  
 387 a 48 hour period (Table 2). While we cannot differentiate the effects of antibiotics on i) increasing  
 388 the susceptibility of an infant to acquisition from another infant, versus ii) endogenous selection  
 389 for resistant bacteria within that infant *Lipsitch and Samore (2002)* in the risk factor models, the  
 390 small estimate for  $\alpha$  relative to  $\beta$  in transmission model 2 suggests that background selection plays  
 391 a relatively small role in transmission compared with person-to-person transmission (Table 3).

392 Breast milk feeding was associated with a reduced risk of colonisation with 3GC-R *K. pneumoniae*  
 393 s.l. though uncertainty was large (Figure 2A). This accords with our understanding of the develop-  
 394 ment of a healthy gut microbiome in the early stages of life, which can be adversely affected by the  
 395 replacement of breast milk by formula *Bäckhed et al. (2015)*, and the protective effect of a diverse  
 396 microbiome that competes against potentially pathogenic bacteria *Langdon et al. (2016)*. In this  
 397 population breastfeeding rates were high (90%) though our simulations showed that dropping the  
 398 proportion to 50% or 25%, as seen in other developing world populations *Lauer et al. (2004)*, could  
 399 increase the proportion becoming colonised during admission by around 5% and 8% respectively.

400 The finding that the oral probiotic *Lactobacillus acidophilus* was not strongly protective against  
 401 acquisition was disappointing in light of earlier results that suggested a possible effect in slowing  
 402 rates of acquisition of ESBL-producing Enterobacteriaceae *Turner et al. (2016)*. This negative result  
 403 was shown by the odds ratio, which was located close to zero in the risk factor models for acquisition  
 404 (Figure 2A) and supported by the forward simulations which showed only a 4% median decrease in  
 405 infant colonisation rates when 100% of infants were prescribed probiotics (Figure 4B). Evidence is still  
 406 limited about the value of probiotics for neonates. One large randomised trial in rural India reported  
 407 a beneficial effect from a symbiotic preparation (combining a probiotic, *Lactobacillus plantarum*, with  
 408 fructooligosaccharide) in preventing sepsis in infants *Panigrahi et al. (2017)*. Another randomised  
 409 trial in pre-term infants in England found no benefit from the probiotic *Bifidobacterium breve* BBG-  
 410 001 in preventing necrotising enterocolitis, blood culture positive sepsis or death before hospital  
 411 discharge *Costeloe et al. (2016)*. There is also evidence that some probiotics given after antibiotic  
 412 consumption can impair and delay the recovery of normal gut flora in humans *Suez et al. (2018)*.

413 An increased number of nurses on the ward was negatively associated with the risk of acquiring  
 414 3GC-R *K. pneumoniae* s.l.. Our simulation studies using a dynamic agent-based model showed  
 415 that increasing the nurse:infant ratio from 1:3 to 1:1 could reduce the number of infants becom-  
 416 ing colonised in both high and low importation settings by about a quarter. These findings are  
 417 consistent with results from large observational studies. For example, *Rogowski et al. (2013)*, in  
 418 a retrospective cohort study in 67 neonatal units in the USA found a strong association between  
 419 neonatal unit understaffing and an increased nosocomial infection rate (where understaffing was  
 420 defined as a nurse-patient ratio below US guidelines for the patient acuity level). Two distinct  
 421 mechanisms might account for such an association. First, an imbalance between workload and  
 422 staffing levels may lead to reduced attention to basic infection control measures such as hand  
 423 hygiene, as has been reported in several studies *Pittet et al. (2006)*. Second, as nurse:patient ratios  
 424 decrease a lower proportion of patient contacts will be cohorted as each nurse will need to contact  
 425 more patients in a shift, substantially increasing the potential for cross-transmission *Archibald et al.*  
 426 *(1997); Austin et al. (1999); Hugonnet et al. (2006)*.

427 From our model estimates, we observed considerable variation in the risk of acquisition per  
 428 patient day, with the median posterior probability varying nearly eight-fold from 0.047 to 0.35,



showing that even neonates within a single hospital ward constitute a highly heterogeneous population. This challenges the assumptions of compartmental models for nosocomial transmission that treat patients as broadly homogeneous in their risk of acquiring drug-resistant bacteria *Grundmann and Hellriegel (2006); van Kleef et al. (2013); de Cellès et al. (2013)*.

The effect of colonisation pressure was not identifiable when we considered the total number of 3GC-R *K. pneumoniae s.l.* isolates on a given day, but became identifiable when we considered transmission by ST. From our analysis of the *Klebsiella* carriage population, we observed three species (*K. pneumoniae*, *K. quasipneumoniae* and *K. variicola*) along with a considerable number (62) of STs. This indicates that *Klebsiella* in carriage did not reflect a single population, but rather repeated introductions of diverse isolates which were then either spread around the ward over a number of weeks or were not transmitted and became extinct.

Comparison of different transmission models strongly supported the inclusion of a colonisation pressure term to account for patient-to-patient transmission. There was also support for a hierarchical model where transmissibility varied by ST, though the STs we estimated to have the highest transmissibility have not been highlighted as dominant in other settings *Wyres and Holt (2016)* suggesting that the ST composition within a region may reflect adaptation to local pressures *Stoesser et al. (2015b)*. The pairwise SNP diversity in two of the major STs (ST45 and ST101) was greater than expected, but the within-host and between-host diversity was similar in both cases, suggesting that transmission within these STs is biologically plausible. Within-host diversity has the potential to confound the reconstruction of transmission networks *Worby et al. (2014); Didelot et al. (2016)*, and our results here highlight the importance of capturing within-host diversity in sequencing studies. While within-host diversity was particularly high in ST45, all acquisitions of this ST occurred within a 19 day window (over a possible four month period when carriage isolates were sequenced) with overlapping colonised patient stays, strongly suggesting that cases were epidemiologically related (Figure 3E). A previous analysis of a subset of the genomic data identified closely related clusters suggestive of transmission, though with a smaller number of pairwise SNPs than we observed *Smit et al. (2018)*. This discrepancy is likely due to differences in methodology as we mapped isolates to ST-specific reference genomes, which results in a greater proportion of the genome being callable compared with mapping very diverse isolates to a single reference. We cannot entirely rule out less parsimonious explanations for the temporal clustering of STs, such as a transient increase of certain STs in the water supply, though the person-to-person transmission route is most strongly supported by our models.

An important strength of the study is that we considered asymptomatic carriage, rather than clinical isolates, and collected detailed patient-level data from those who became colonised as well as from those who remained uncolonised. This allowed us to quantify the factors driving the epidemic, which would not have been possible if we had considered only clinical isolates. Also, by using a prospective design rather than a reactive exploration of an outbreak or unusual cluster of cases our findings should be more representative of typical patterns of transmission. The use of an inferential approach that accounted for the interval-censored nature of the data and the use of whole genome sequencing were also key factors in developing a quantitative understanding of the transmission dynamics.

Our study has limitations. We sequenced isolates of 3GC-R *K. pneumoniae s.l.* from a four month period (rather than the full 12 months of the carriage study) due to resource constraints. Such constraints also prevented us from sampling mothers and other family members of infants, and healthcare workers in close contact with infants. Greater density of within-host sampling *Stoesser et al. (2015a); Wymant et al. (2017); Lees et al. (2019)*, and inclusion of long-read sequencing to investigate plasmid transmission *Conlan et al. (2014)* would have also provided a more complete epidemiological picture. Although our analysis accounts for patient heterogeneity and interval censoring, it assumes that a positive culture taken within 48 hours of admission indicates that the patient was colonised on admission, and that cultures had 100% sensitivity for detecting carriage. Such restrictions could all be relaxed using a data augmentation framework *Cooper et al. (2008)*,



but validated implementations of such an approach allowing inclusion of an arbitrary number of covariates are not currently available. Antibiotics were not prescribed randomly, but according to the clinical judgement of the clinicians and in response to the pathology of the patient. It is therefore possible that the effects of antibiotics in the model estimates may be confounded by the clinical severity of the infant. We did attempt to mitigate this by including a covariate for clinical severity, which links to the use of invasive devices and we note that, with the exception of imipenem, all antibiotics show similar effects to each other on the risk of first acquisition.

In summary, this study provides strong evidence for person-to-person within ward transmission of 3GC-R *K. pneumoniae* s.l. mediated by healthcare workers, estimates key epidemiological parameters, quantifies the role of different antibiotics in driving the epidemic, and highlights interventions with the potential to contribute to control efforts.

## Methods and Materials

### Epidemiological and Microbiological Data

We used data collected prospectively from a neonatal intensive care unit in a children's hospital in Siem Reap, Cambodia between September 2013 (when the neonatal unit newly opened) and September 2014. The study protocol required rectal swabs to be taken within 48 hours of admission and subsequently every two to three days. We assumed that any patient testing positive on their first swab taken within 48 hours of admission was positive on arrival, and therefore excluded from the analysis. Patients who had their first swab taken >48 hours from admission but tested negative were included in the analysis, while those that tested positive were omitted. Infants that were re-admitted to the ward and had been colonised prior to first discharge were assumed to still be colonised and were thus excluded. Details of the microbiological treatment of rectal swabs, including resistance assays, have been published previously [Turner et al. \(2016\)](#).

### Probability of Acquisition

We estimated the daily probability of colonisation with 3GC-R *K. pneumoniae* s.l. using a discrete time model with time steps of one day. Each day in the ward a previously uncolonised patient can become colonised. As rectal swabs are not taken on every day of a patient's stay the outcome is interval censored: we know that a negative swab followed by a positive swab indicates that a patient became colonised on some day between the two swabs, but not on which day. If the probability of becoming colonised on day  $i$  for patient  $j$  is  $p_{ij}$ , given the patient is uncolonised at the start of the day, then the probability of remaining uncolonised is  $(1-p_{ij})$ . In interval  $k$  for patient  $j$  consisting of  $N_{kj}$  days, then the probability of remaining uncolonised is:

$$\prod_{i=1}^{N_{kj}} (1 - p_{ij})$$

Therefore the probability of becoming colonised  $v_{kj}$  is the complement:

$$v_{kj} = 1 - \prod_{i=1}^{N_{kj}} (1 - p_{ij})$$

The outcome for patient  $j$  in interval  $k$ ,  $Y_{kj} \in \{0,1\}$ , as the patient either becomes colonised (1) or remains uncolonised (0) with 3GC-R *Klebsiella*. Therefore the likelihood is given by:

$$Y_{kj} \sim \text{Bernoulli}(v_{kj})$$

### Risk Factor Models for Carriage Acquisition

The daily probability of becoming colonised ( $p_{ij}$ ) is related by the logit link function to a linear function of covariates:

$$\pi_{ij} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \dots$$

$$p_{ij} = \frac{\exp(\pi_{ij})}{\exp(\pi_{ij}) + 1}$$

Where  $x_1, x_2, x_3 \dots$  is a vector of predictors (data) and  $\beta_1, \beta_2, \beta_3 \dots$  is a vector of slopes (parameters) that are to be estimated. The intercept  $\alpha$  can be a single parameter, or permitted to vary over  $m$  periods of time in a random effects model. When using such a random effects model  $\alpha_m$  was assumed to be normally distributed, with a mean  $\mu$  and standard deviation  $\sigma$  which are themselves parameters with their own prior distributions. The prior distributions used in the analysis are shown in Table 2. Results from models with alternative prior distributions are shown in Figure 2- figure supplement 2.

$$\alpha_m \sim \text{normal}(\mu, \sigma)$$

Fourteen standard covariates included in all risk factor models were: exposure to the six most common combinations of antibiotics, taken intravenously unless otherwise stated (ampicillin, ampicillin + gentamicin, cloxacillin (oral), ceftriaxone, cloxacillin + gentamicin, and imipenem) within the past 48 or 96 hours, whether the infant was breast fed, if the infant received an oral probiotic on entry (*Lactobacillus acidophilus*) for prevention of necrotising enterocolitis, sex, born prematurely (before the 37th week of pregnancy), severity (defined as either i. requiring ventilation, ii. requiring continuous positive airway pressure or iii. requiring intubation), and if already colonised with 3GC-R *E. coli*. These explanatory variables were treated as binary (0/1). We also included the age in days on first admission to the NU, and the daily number of nurses on the ward. An additional covariate included in one of the risk factor models (see below) was a term for colonisation pressure, which is an integer value representing the number of individuals known to be colonised with 3GC-R *K. pneumoniae s.l.* on that day. Covariates were recorded for every day the infant was present in the neonatal unit and data were treated as complete. We considered the following models:

- A. Single intercept with standard covariates (14). Exposure to antibiotics considered if taken within the past 96 hours.
- B. Single intercept with standard covariates (14). Exposure to antibiotics considered if taken within the past 48 hours.
- C. Single intercept with standard covariates (14) plus an additional covariate for colonisation pressure. Exposure to antibiotics considered if taken within the past 96 hours.
- D. Variable intercept by study month that uses partial pooling (hierarchical model). Standard covariates (14) and exposure to antibiotics considered if taken within the past 96 hours.

### Rectal Swab / Culture Sensitivity

Swab sensitivity was estimated from the number of negative swabs following a positive swab, i) under the assumption that all negative results following a positive swab were false negatives and ii) under the assumption that three or more consecutive negative swabs following a positive represented a true decolonisation event. Posterior distributions of the false negative rate and swab sensitivity were estimated using a conjugate beta prior;  $\beta_1(\alpha=1, \beta=7)$  **Bolker (2008)**.

### Pathogen Sequencing and Bioinformatics

We whole-genome sequenced 317 cultured isolates identified morphologically as 3GC-R *K. pneumoniae s.l.* from i) rectal swabs from all colonised patients within a four month period of the study and ii) twice weekly swabs from 7 environmental surfaces around the ward (6 sinks and 1 computer keyboard) within the same time frame. Sequencing was performed with the Illumina HiSeq 2500 platform, producing 150 base-pair paired-end reads. The reads were trimmed for adapter sequence using TrimGalore v0.4.4 before assembly with Unicycler v0.4.5 **Wick et al.**

(2017), contigs <1 kilobase were discarded. Distances between assemblies were calculated using mash v1.1 *Ondov et al. (2016)* and a phylogeny constructed with mashtree v0.33. Sequence types (STs) were identified using Kleborate 0.2.0 *Wyres et al. (2016)*. Variant calling was performed within STs by mapping reads to ST consensus reference genomes (5.32 Mbp) with SMALT (v.0.7.6) <https://www.sanger.ac.uk/science/tools/smalt-0> (parameters -x -y 0.85 -r 1 -j 100 -i 800), before sorting and removing unpaired mate reads and technical duplicates from binary alignment files with samtools (v.1.8). Single nucleotide variants (SNPs) were called by piping output from samtools mpileup into bcftools (v.1.8) and were conservatively filtered to remove SNPs within 50bp of indels, with a read depth <10x and >200x, a mapping quality score <30 or read quality score <100. Repetitive regions were identified with nucmer (v.3.1), phage regions with PHASTER *Arndt et al. (2016)*, and recombination with ClonalFrameML (v.1.11) *Didelot and Wilson (2015)*, and these regions subsequently masked from SNP calling.

### Transmission Models Incorporating Sequence Type Data

We assessed within-ward transmission of 3GC-R *K. pneumoniae* s.l. STs under the assumption that individuals remain colonised with a given ST for the duration of their stay until discharge *Birgand et al. (2013)*. We fitted data to five linear transmission models.

1. Where the daily risk of acquiring ST  $c$  is constant and unlinked to the number of other infants colonised with ST  $c$  on that day (intercept only).
2. A constant term plus a covariate for colonisation pressure ( $\beta$ ), where the explanatory variable ( $n_{ic}$ ) is the number of other patients colonised with ST  $c$  in the neonatal unit on day  $i$ .
3. As (2) with an additional term for the contribution of environmental contamination to transmission ( $\gamma$ ), whereby the number of cases with ST  $c$  on an earlier day  $i'$  ( $n_{i'c}$ ) are assumed to leave some trace in the environment that decays exponentially over time at a rate given by  $\lambda$ .  $N$  gives the total number of days.
4. A hierarchical version of (2) which permits the transmission parameter  $\beta$  to vary by ST  $c$ .
5. A hierarchical version of (2) which permits the intercept  $\alpha$  to vary by ST  $c$ .

The probability  $p$  of colonisation for individual  $j$  on day  $i$  with ST  $c$  for the respective models are:

$$p_{ijc} = \alpha \quad \text{(Transmission Model 1)}$$

$$p_{ijc} = \alpha + \beta n_{ic} \quad \text{(Transmission Model 2)}$$

$$p_{ijc} = \alpha + \beta n_{ic} + \gamma \sum_{i'=0}^{i-1} n_{i'c} e^{-\lambda(i-i')} \quad \text{(Transmission Model 3)}$$

$$p_{ijc} = \alpha + \beta_c n_{ic} \quad \text{(Transmission Model 4)}$$

$$p_{ijc} = \alpha_c + \beta n_{ic} \quad \text{(Transmission Model 5)}$$

We opted to fit transmission models on the linear, rather than logistic, scale as we consider a linear increase in the force of infection with the number of colonised infants to be a more realistic assumption than the, initially, exponential increase that results from a logistic transformation. Beta distributions were used as priors, in the case of the hierarchical transmission models (4 and 5) the scale and location parameters are themselves hyper-parameters with their own priors. Studies have shown the half life of carbapenem resistant *K. pneumoniae* inoculated onto hospital surfaces to be around 12 hours *Weber et al. (2015)*, this corresponds to  $\lambda=1.39$  in our model. We therefore gave  $\lambda$  in transmission model 3 a more informative prior, normal( $\mu=1$ ,  $\sigma=2$ ), to give weight to biologically plausible estimates. Prior distributions for all transmission models are shown in Table 3. Results from a less informative prior distribution for  $\lambda$  are shown in Figure 3- figure supplement 2.

### Statistical Model Fitting

We fitted the statistical models using Hamiltonian Markov Chain Monte Carlo in Stan (version 2.17.3) within the R environment (v. 3.4.3). Prior distributions were selected to be weakly informative

normal distributions for the risk-factor models, and beta distributions for the transmission models **McElreath (2018)**, see Tables 2 and 3 for all prior distributions. Chains were run for a varying number of iterations depending on the number of parameters to estimate, though with a minimum of 12,000 iterations over four chains, including burn-in. The Gelman-Rubin statistic ( $\hat{R}$ ) was used as a diagnostic, where values  $<1.01$  indicate chains have converged and additionally posterior chains were visually inspected for convergence. Effective sample sizes (ESS) for both the centre of the posterior distribution (bulk) and the ends of the distribution (tail) was ensured to be  $>400$  **Vehtari et al. (2019)**. Model comparison was performed with widely applicable information criterion (WAIC) **Vehtari et al. (2017)**. We used 95% credible intervals (CrIs) as a measure of uncertainty around posterior parameter distributions and posterior medians as the central estimate. Uncertainty in parameter estimates is represented by the posterior distributions, and the estimated probability that a particular parameter is within a certain range is given by the area under the curve of that parameter's marginal posterior distribution within that range.

## Agent Based Forward Simulations

We forward simulated the impact of interventions using an agent-based model implemented in Python (v. 2.7.15) and hosted at <https://github.com/tc13/ward-infection-ABM>. In brief, we model in discrete time steps a ward containing a fixed number of beds and where patients sample a length of stay and colonisation status on entry. Simulations to estimate the ward reproduction number  $R_A$  introduced a single colonised individual into a full ward of susceptible patients. The probability of any uncolonised patient acquiring resistant *Klebsiella* from the index patient on day 1 is  $p$ , where  $p$  is sampled from a beta distribution with shape hyper-parameters  $\alpha_s$  and  $\beta_s$  from transmission model 4 (Table 3). This probability changes on subsequent days based on the number of other infants that become colonised and start transmission, and the lengths of stay. The simulation ends when the index patient is discharged.

For simulations that explored the impact of interventions, the effect of each of the covariates of interest (probiotic consumption, breast milk feeding and number of nurses on the ward) were obtained from risk factor model A (Table 2 and Figure 2A). The marginal effects of each risk factor was transformed into a log-odds ratio and used to modify the daily risk of acquisition sampled from a beta distribution with shape hyper-parameters  $\alpha_s$  and  $\beta_s$ . For  $n_c$  colonised patients on the ward on day  $j$ , the probability of an uncolonised patient with covariate  $k$  acquiring 3GC-R *K. pneumoniae* s.l. is  $1-(1-p_k)^{n_{cj}}$ . The simulation ends after 365 days. To reduce variability between model runs for estimation of  $R_A$  and the interventions, simulations were run 100 times with each posterior parameter sample and the mean outcome value obtained. As we used 2000 posterior samples from each parameter estimated by model fitting, this resulted in a total of 200,000 model runs for each simulated scenario.

## Data and Code Availability

Code for reproducing the statistical model fitting and anonymised patient data are available at <https://github.com/tc13/transmission-dynamics-klebsiella>. The code for the agent based model and parameter values for forward simulations are available at <https://github.com/tc13/ward-infection-ABM>. Short read sequence data is available from NCBI under accession numbers PRJNA395864 and PRJEB24970. Genomic data can be visualised in the Microreact project <https://microreact.org/project/BV05TIXKU>.

## Acknowledgments

We thank the staff and patients' families at the Angkor Hospital for Children for their participation in the study. We also thank Anastasia Hernandez Koutoucheva and Patrick Musicha for their comments on the manuscript.

## Funding Statement

BSC and TC were supported by The Medical Research Council and Department for International Development (grant number MR/K006924/1). BSC works within the Wellcome Trust Major Overseas Programme in SE Asia (grant number 106698/Z/14/Z). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## References

- Adler JL**, Shulman JA, Terry PM, Feldman DB, Skaliy P. Nosocomial colonization with kanamycin-resistant *Klebsiella pneumoniae*, types 2 and 11, in a premature nursery. *The Journal of pediatrics*. 1970; 77(3):376–385.
- Anderson RM**, May RM. *Infectious diseases of humans: dynamics and control*. Oxford university press; 1992.
- Andriatahina T**, Randrianirina F, Hariniana ER, Talarmin A, Raobijaona H, Buisson Y, Richard V. High prevalence of fecal carriage of extended-spectrum  $\beta$ -lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* in a pediatric unit in Madagascar. *BMC infectious diseases*. 2010; 10(1):204.
- Archibald LK**, Manning ML, Bell LM, Banerjee S, Jarvis WR. Patient density, nurse-to-patient ratio and nosocomial infection risk in a pediatric cardiac intensive care unit. *The Pediatric infectious disease journal*. 1997; 16(11):1045–1048.
- Arndt D**, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic acids research*. 2016; 44(W1):W16–W21.
- Austin DJ**, Bonten MJ, Weinstein RA, Slaughter S, Anderson RM. Vancomycin-resistant enterococci in intensive-care hospital settings: transmission dynamics, persistence, and the impact of infection control programs. *Proceedings of the National Academy of Sciences*. 1999; 96(12):6908–6913.
- Bäckhed F**, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H, et al. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell host & microbe*. 2015; 17(5):690–703.
- Baker S**, Thomson N, Weill FX, Holt KE. Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. *Science*. 2018; 360(6390):733–738.
- Birgand G**, Armand-Lefevre L, Lolom I, Ruppe E, Andremont A, Lucet JC. Duration of colonization by extended-spectrum  $\beta$ -lactamase-producing Enterobacteriaceae after hospital discharge. *American journal of infection control*. 2013; 41(5):443–447.
- Bolker BM**. *Ecological models and data in R*. Princeton University Press; 2008.
- Bonneault M**, Andrianoelina VH, Herindrainy P, Rabenandrasana MAN, Garin B, Breurec S, Delarocque-Astagneau E, Guillemot D, Andrianirina ZZ, Collard JM, et al. Transmission Routes of Extended-Spectrum Beta-Lactamase-Producing Enterobacteriaceae in a Neonatology Ward in Madagascar. *The American journal of tropical medicine and hygiene*. 2019; 100(6):1355–1362.
- Bonten MJ**. Colonization pressure: a critical parameter in the epidemiology of antibiotic-resistant bacteria. *Critical Care*. 2012; 16(4):142.
- de Cellès MD**, Zahar JR, Abadie V, Guillemot D. Limits of patient isolation measures to control extended-spectrum beta-lactamase-producing Enterobacteriaceae: model-based analysis of clinical data in a pediatric ward. *BMC infectious diseases*. 2013; 13(1):187.
- Ciccolini M**, Donker T, Köck R, Mielke M, Hendrix R, Jurke A, Rahamat-Langendoen J, Becker K, Niesters HGM, Grundmann H, et al. Infection prevention in a connected world: The case for a regional approach. *International Journal of Medical Microbiology*. 2013; 303(6-7):380–387.
- Clarivet B**, Grau D, Jumas-Bilak E, Jean-Pierre H, Pantel A, Parer S, Lotthé A. Persisting transmission of carbapenemase-producing *Klebsiella pneumoniae* due to an environmental reservoir in a university hospital, France, 2012 to 2014. *Eurosurveillance*. 2016; 21(17):30213.
- Conlan S**, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, Snitkin ES, Clark TA, Luong K, Song Y, et al. Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. *Science translational medicine*. 2014; 6(254):254ra126–254ra126.

- 693 **Cooper BS**, Medley GF, Bradley SJ, Scott GM. An augmented data method for the analysis of nosocomial infection  
694 data. *American journal of epidemiology*. 2008; 168(5):548–557.
- 695 **Costeloe K**, Hardy P, Juszczak E, Wilks M, Millar MR, et al. Bifidobacterium breve BBG-001 in very preterm infants:  
696 a randomised controlled phase 3 trial. *The Lancet*. 2016; 387(10019):649–660.
- 697 **Decraene V**, Phan H, George R, Wyllie D, Akinremi O, Aiken Z, Cleary P, Dodgson A, Pankhurst L, Crook D, et al. A  
698 Large, Refractory Nosocomial Outbreak of *Klebsiella pneumoniae* Carbapenemase-Producing *Escherichia coli*  
699 Demonstrates Carbapenemase Gene Outbreaks Involving Sink Sites Require Novel Approaches to Infection  
700 Control. *Antimicrobial agents and chemotherapy*. 2018; 62(12):e01689–18.
- 701 **Didelot X**, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. *Nature*  
702 *Reviews Microbiology*. 2016; 14(3):150.
- 703 **Didelot X**, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS*  
704 *computational biology*. 2015; 11(2):e1004041.
- 705 **Dondorp AM**, Limmathurotsakul D, Ashley EA. What's wrong in the control of antimicrobial resistance in critically  
706 ill patients from low-and middle-income countries? *Intensive care medicine*. 2018; 44(1):79–82.
- 707 **Donker T**, Wallinga J, Slack R, Grundmann H. Hospital networks and the dispersal of hospital-acquired pathogens  
708 by patient transfer. *PLoS One*. 2012; 7(4):e35002.
- 709 **Duval A**, Obadia T, Boëlle PY, Fleury E, Herrmann JL, Guillemot D, Temime L, Opatowski L, Group IBS, et al. Close  
710 proximity interactions support transmission of ESBL-K. *pneumoniae* but not ESBL-E. *coli* in healthcare settings.  
711 *PLoS computational biology*. 2019; 15(5):e1006496.
- 712 **Forrester M**, Pettitt AN. Use of stochastic epidemic modeling to quantify transmission rates of colonization  
713 with methicillin-resistant *Staphylococcus aureus* in an intensive care unit. *Infection Control & Hospital*  
714 *Epidemiology*. 2005; 26(7):598–606.
- 715 **Founou RC**, Founou LL, Allam M, Ismail A, Essack SY. Whole Genome Sequencing of Extended Spectrum  
716  $\beta$ -lactamase (ESBL)-producing *Klebsiella pneumoniae* Isolated from Hospitalized Patients in KwaZulu-Natal,  
717 South Africa. *Scientific Reports*. 2019; 9(1):6266.
- 718 **Fox-Lewis A**, Takata J, Miliya T, Lubell Y, Soeng S, Sar P, Rith K, McKellar G, Wuthiekanun V, McGonagle E,  
719 et al. Antimicrobial resistance in invasive bacterial infections in hospitalized children, Cambodia, 2007–2016.  
720 *Emerging infectious diseases*. 2018; 24(5):841.
- 721 **Goldmann DA**. Bacterial colonization and infection in the neonate. *The American journal of medicine*. 1981;  
722 70(2):417–422.
- 723 **Gorrie CL**, Mirčeta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, Pratt NF, Garlick JS, Watson KM, Pilcher  
724 DV, et al. Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care  
725 patients. *Clinical Infectious Diseases*. 2017; 65(2):208–215.
- 726 **Gorrie CL**, Mirceta M, Wick RR, Judd LM, Wyres KL, Thomson NR, Strugnell RA, Pratt NF, Garlick JS, Watson KM,  
727 et al. Antimicrobial-resistant *klebsiella pneumoniae* carriage and infection in specialized geriatric care wards  
728 linked to acquisition in the referring hospital. *Clinical Infectious Diseases*. 2018; 67(2):161–170.
- 729 **Grundmann H**, Hellriegel B. Mathematical modelling: a tool for hospital infection control. *The Lancet infectious*  
730 *diseases*. 2006; 6(1):39–45.
- 731 **Gurieva T**, Dautzenberg MJD, Gniadkowski M, Derde LPG, Bonten MJM, Bootsma MCJ. The transmissibility of  
732 antibiotic-resistant Enterobacteriaceae in Intensive Care Units. *Clinical Infectious Diseases*. 2017; 66(4):489–  
733 493.
- 734 **Haller S**, Eller C, Hermes J, Kaase M, Steglich M, Radonić A, Dabrowski PW, Nitsche A, Pfeifer Y, Werner G,  
735 et al. What caused the outbreak of ESBL-producing *Klebsiella pneumoniae* in a neonatal intensive care  
736 unit, Germany 2009 to 2012? Reconstructing transmission with epidemiological analysis and whole-genome  
737 sequencing. *BMJ open*. 2015; 5(5):e007397.
- 738 **Holt KE**, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, Jenney A, Connor TR, Hsu LY, Severin J,  
739 et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella*  
740 *pneumoniae*, an urgent threat to public health. *Proceedings of the National Academy of Sciences*. 2015;  
741 112(27):E3574–E3581.

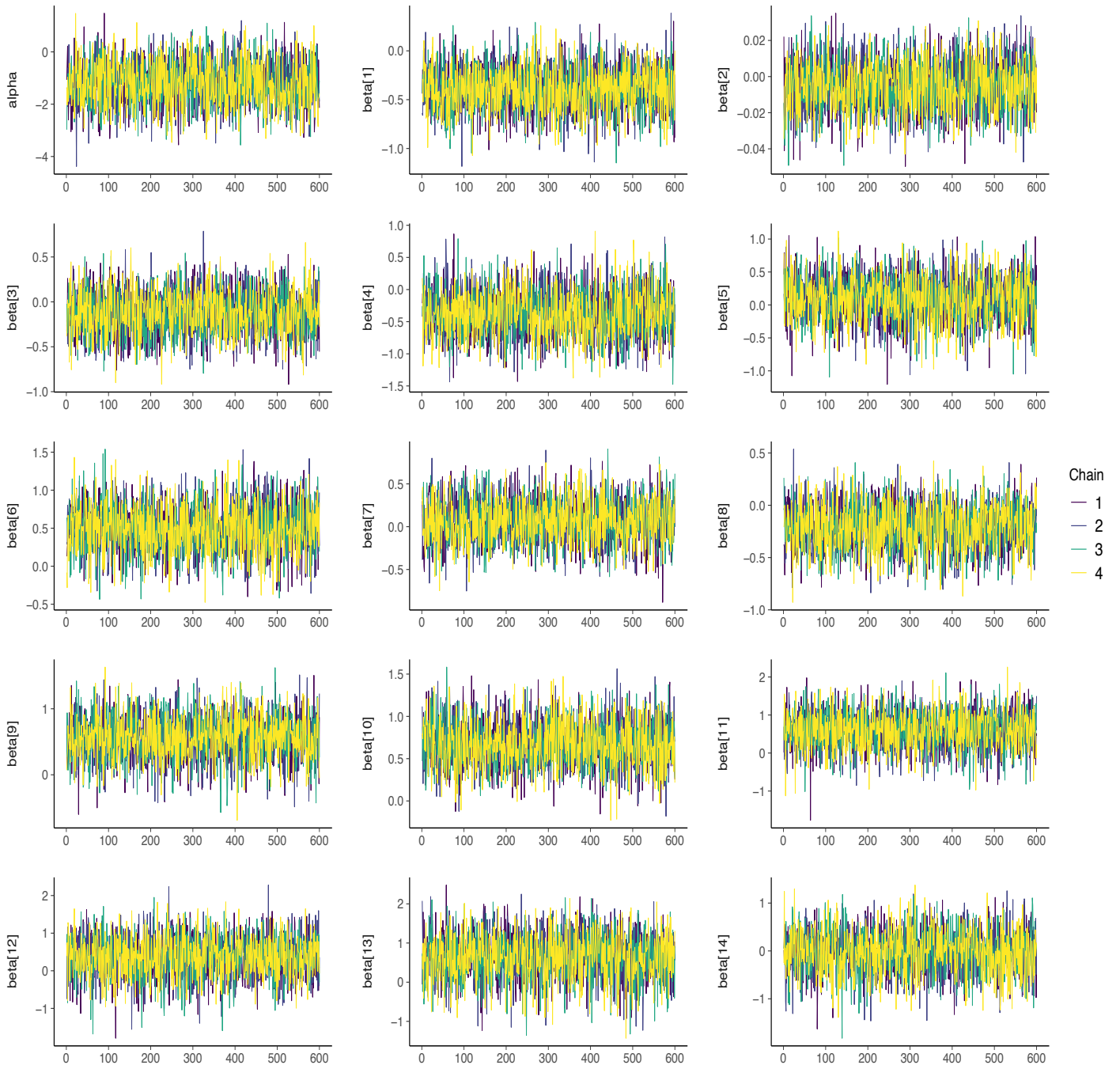


- 742 **Hugonnet S**, Harbarth S, Sax H, Duncan RA, Pittet D. Nursing resources: a major determinant of nosocomial  
743 infection? *Archives of pediatrics & adolescent medicine*. 2006; 160(8):832.
- 744 **Kessner DM**, Lepper MH. Epidemiologic studies of gram-negative bacilli in the hospital and community.  
745 *American journal of epidemiology*. 1967; 85(1):45–60.
- 746 **van Kleef E**, Robotham JV, Jit M, Deeny SR, Edmunds WJ. Modelling the transmission of healthcare associated  
747 infections: a systematic review. *BMC infectious diseases*. 2013; 13(1):294.
- 748 **Kresky B**. Control of GramNegative Bacilli in a Hospital Nursery. *American Journal of Diseases of Children*. 1964;  
749 107(4):363–369.
- 750 **Langdon A**, Crook N, Dantas G. The effects of antibiotics on the microbiome throughout development and  
751 alternative approaches for therapeutic modulation. *Genome medicine*. 2016; 8(1):39.
- 752 **Lauer JA**, Betrán AP, Victora CG, De Onís M, Barros AJD. Breastfeeding patterns and exposure to suboptimal  
753 breastfeeding among children in developing countries: review and analysis of nationally representative  
754 surveys. *BMC medicine*. 2004; 2(1):26.
- 755 **Lees JA**, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, Corander J, Bentley SD, Croucher NJ. Fast and  
756 flexible bacterial genomic epidemiology with PopPUNK. *Genome research*. 2019; 29(2):304–316.
- 757 **Lim C**, Takahashi E, Hongsuwan M, Wuthiekanun V, Thamlikitkul V, Hinjoy S, Day NPJ, Peacock SJ, Limmathurot-  
758 sakul D. Epidemiology and burden of multidrug-resistant bacterial infection in a developing country. *Elife*.  
759 2016; 5:e18082.
- 760 **Lipsitch M**, Bergstrom CT, Levin BR. The epidemiology of antibiotic resistance in hospitals: paradoxes and  
761 prescriptions. *Proceedings of the National Academy of Sciences*. 2000; 97(4):1938–1943.
- 762 **Lipsitch M**, Samore MH. Antimicrobial use and antimicrobial resistance: a population perspective. *Emerging*  
763 *infectious diseases*. 2002; 8(4):347.
- 764 **Martin RM**, Cao J, Brisse S, Passet V, Wu W, Zhao L, Malani PN, Rao K, Bachman MA. Molecular epidemiology of  
765 colonizing and infecting isolates of *Klebsiella pneumoniae*. *mSphere*. 2016; 1(5):e00261–16.
- 766 **Mathers AJ**, Vegesana K, German Mesner I, Barry KE, Pannone A, Baumann J, Crook DW, Stoesser N, Kotay S,  
767 Carroll J, et al. Intensive care unit wastewater interventions to prevent transmission of multispecies *Klebsiella*  
768 *pneumoniae* carbapenemase-producing organisms. *Clinical Infectious Diseases*. 2018; 67(2):171–178.
- 769 **McElreath R**. Statistical Rethinking: A Bayesian Course with Examples in R and Stan. CRC Press; 2018.
- 770 **Mertz JJ**, Scharer L, McClement JH. A hospital outbreak of *Klebsiella pneumoniae* from inhalation therapy with  
771 contaminated aerosol solutions. *American Review of Respiratory Disease*. 1967; 95(3):454–460.
- 772 **Montgomerie JZ**. Epidemiology of *Klebsiella* and hospital-associated infections. *Reviews of infectious diseases*.  
773 1979; 1(5):736–753.
- 774 **Musicha P**, Cornick JE, Bar-Zeev N, French N, Masesa C, Denis B, Kennedy N, Mallewa J, Gordon MA, Msefula CL,  
775 et al. Trends in antimicrobial resistance in bloodstream infection isolates at a large urban hospital in Malawi  
776 (1998–2016): a surveillance study. *The Lancet infectious diseases*. 2017; 17(10):1042–1052.
- 777 **Nordmann P**, Naas T, Poirel L. Global spread of carbapenemase-producing Enterobacteriaceae. *Emerging*  
778 *infectious diseases*. 2011; 17(10):1791.
- 779 **Ondov BD**, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. Mash: fast genome and  
780 metagenome distance estimation using MinHash. *Genome biology*. 2016; 17(1):132.
- 781 **Palleja A**, Mikkelsen KH, Forslund SK, Kashani A, Allin KH, Nielsen T, Hansen TH, Liang S, Feng Q, Zhang C,  
782 et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nature microbiology*. 2018;  
783 3(11):1255.
- 784 **Panigrahi P**, Parida S, Nanda NC, Satpathy R, Pradhan L, Chandel DS, Baccaglini L, Mohapatra A, Mohapatra SS,  
785 Misra PR, et al. A randomized synbiotic trial to prevent sepsis among infants in rural India. *Nature*. 2017;  
786 548(7668):407.
- 787 **Peleg AY**, Hooper DC. Hospital-acquired infections due to gram-negative bacteria. *New England Journal of*  
788 *Medicine*. 2010; 362(19):1804–1813.

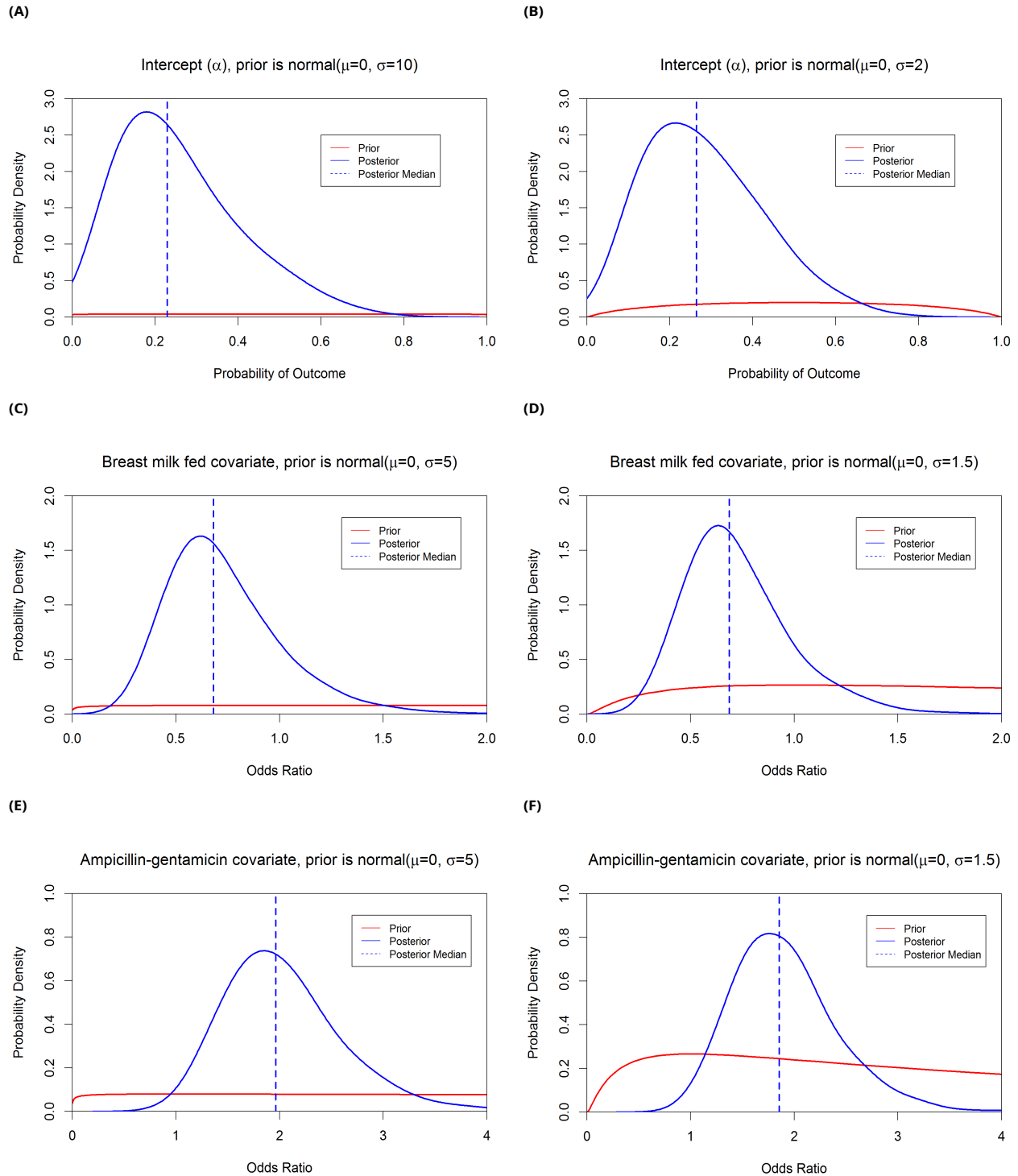
- 789 **Pittet D**, Allegranzi B, Sax H, Dharan S, Pessoa-Silva CL, Donaldson L, Boyce JM, et al. Evidence-based model for  
790 hand transmission during patient care and the role of improved practices. *The Lancet infectious diseases*.  
791 2006; 6(10):641–652.
- 792 **Roberts T**, Limmathurotsakul D, Turner P, Day NP, Vandepitte WP, Cooper BS. Antimicrobial-resistant Gram-  
793 negative colonisation in infants from a NICU in Thailand. *Journal of Hospital Infection*. 2019; .
- 794 **Rogowski JA**, Staiger D, Patrick T, Horbar J, Kenny M, Lake ET. Nurse staffing and NICU infection rates. *JAMA*  
795 *pediatrics*. 2013; 167(5):444–450.
- 796 **Schwaber MJ**, Carmeli Y. Carbapenem-resistant Enterobacteriaceae: a potential threat. *JAMA*. 2008;  
797 300(24):2911–2913.
- 798 **Selden R**, Lee S, Wang WLL, Bennett JV, Eickhoff TC. Nosocomial *Klebsiella* infections: intestinal colonization as  
799 a reservoir. *Annals of Internal Medicine*. 1971; 74(5):657–664.
- 800 **Smit PW**, Stoesser N, Pol S, van Kleef E, Oonsivilai M, Tan P, Neou L, Turner C, Turner P, Cooper BS. Transmission  
801 dynamics of hyper-endemic multi-drug resistant *Klebsiella pneumoniae* in a Southeast Asian neonatal unit: a  
802 longitudinal study with whole genome sequencing. *Frontiers in microbiology*. 2018; 9:1197.
- 803 **Snitkin ES**, Won S, Pirani A, Lapp Z, Weinstein RA, Lolans K, Hayden MK. Integrated genomic and interfacility  
804 patient-transfer data reveal the transmission pathways of multidrug-resistant *Klebsiella pneumoniae* in a  
805 regional outbreak. *Science translational medicine*. 2017; 9(417):eaan0093.
- 806 **Snitkin ES**, Zelazny AM, Thomas PJ, Stock F, Henderson DK, Palmore TN, Segre JA, Program NISCCS, et al. Tracking  
807 a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Science*  
808 *translational medicine*. 2012; 4(148):148ra116–148ra116.
- 809 **Stoesser N**, Sheppard AE, Moore C, Golubchik T, Parry C, Nget P, Saroeun M, Day N, Giess A, Johnson J, et al.  
810 Extensive within-host diversity in fecally carried extended-spectrum-beta-lactamase-producing *Escherichia*  
811 *coli* isolates: implications for transmission analyses. *Journal of clinical microbiology*. 2015; 53(7):2122–2131.
- 812 **Stoesser N**, Xayaheuang S, Vongsouvath M, Phommason K, Elliott I, del Ojo Elias C, Crook DW, Newton PN,  
813 Buisson Y, Lee SJ, Dance DAB. Colonization with Enterobacteriaceae producing ESBLs in children attending  
814 pre-school childcare facilities in the Lao People's Democratic Republic. *Journal of Antimicrobial Chemotherapy*.  
815 2015; 70(6):1893–1897.
- 816 **Suez J**, Zmora N, Zilberman-Schapira G, Mor U, Dori-Bachash M, Bashirdes S, Zur M, Regev-Lehavi D, Brik  
817 RBZ, Federici S, et al. Post-antibiotic gut mucosal microbiome reconstitution is impaired by probiotics and  
818 improved by autologous FMT. *Cell*. 2018; 174(6):1406–1423.
- 819 **Tedijanto C**, Olesen SW, Grad YH, Lipsitch M. Estimating the proportion of bystander selection for antibiotic  
820 resistance among potentially pathogenic bacterial flora. *Proceedings of the National Academy of Sciences*.  
821 2018; 115(51):E11988–E11995.
- 822 **Theuretzbacher U**. Global antimicrobial resistance in Gram-negative pathogens and clinical need. *Current*  
823 *opinion in microbiology*. 2017; 39:106–112.
- 824 **Turner P**, Pol S, Soeng S, Sar P, Neou L, Chea P, Day NPJ, Cooper BS, Turner C. High prevalence of antimicrobial-  
825 resistant Gram-negative colonization in hospitalized Cambodian infants. *The Pediatric infectious disease*  
826 *journal*. 2016; 35(8):856.
- 827 **Vehtari A**, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and  
828 WAIC. *Statistics and Computing*. 2017; 27:1413–1432. doi: 10.1007/s11222-016-9696-4.
- 829 **Vehtari A**, Gelman A, Simpson D, Carpenter B, Bürkner PC. Rank-normalization, folding, and localization: An  
830 improved R-hat for assessing convergence of MCMC. *arXiv preprint arXiv:190308008*. 2019; .
- 831 **Weber DJ**, Rutala WA, Kanamori H, Gergen MF, Sickbert-Bennett EE. Carbapenem-resistant Enterobacteriaceae:  
832 frequency of hospital room contamination and survival on various inoculated surfaces. *infection control &*  
833 *hospital epidemiology*. 2015; 36(5):590–593.
- 834 **Weil AJ**, Ramchand S, Arias ME. Nosocomial infection with *Klebsiella* type 25. *New England Journal of Medicine*.  
835 1966; 275(1):17–22.
- 836 **Wick RR**, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long  
837 sequencing reads. *PLoS computational biology*. 2017; 13(6):e1005595.

- 838 **Worby CJ**, Lipsitch M, Hanage WP. Within-host bacterial diversity hinders accurate reconstruction of transmission  
839 networks from genomic distance data. *PLoS computational biology*. 2014; 10(3):e1003549.
- 840 **World Health Organization**. Antimicrobial resistance: global report on surveillance. World Health Organization;  
841 2014.
- 842 **World Health Organization**. Global priority list of antibiotic-resistant bacteria to guide research, discovery,  
843 and development of new antibiotics. Geneva: World Health Organization. 2017; .
- 844 **Wymant C**, Hall M, Ratmann O, Bonsall D, Golubchik T, de Cesare M, Gall A, Cornelissen M, Fraser C. PHY-  
845 LOSCANNER: inferring transmission from within-and between-host pathogen genetic diversity. *Molecular*  
846 *biology and evolution*. 2017; 35(3):719–733.
- 847 **Wyres KL**, Holt KE. *Klebsiella pneumoniae* population genomics and antimicrobial-resistant clones. *Trends in*  
848 *microbiology*. 2016; 24(12):944–956.
- 849 **Wyres KL**, Wick RR, Gorrie C, Jenney A, Follador R, Thomson NR, Holt KE. Identification of *Klebsiella* capsule  
850 synthesis loci from whole genome data. *Microbial genomics*. 2016; 2(12).
- 851 **Yow EM**. Clinical significance of rising incidence of infections due to gram-negative bacilli. *Postgraduate*  
852 *medicine*. 1955; 17(5):413–419.
- 853 **Zaidi AKM**, Huskins WC, Thaver D, Bhutta ZA, Abbas Z, Goldmann DA. Hospital-acquired neonatal infections in  
854 developing countries. *The Lancet*. 2005; 365(9465):1175–1188.
- 855 **Zellweger RM**, Carrique-Mas J, Limmathurotsakul D, Day NPJ, Thwaites GE, Baker S, Network SAAR. A current  
856 perspective on antimicrobial resistance in Southeast Asia. *Journal of Antimicrobial Chemotherapy*. 2017;  
857 72(11):2963–2972.

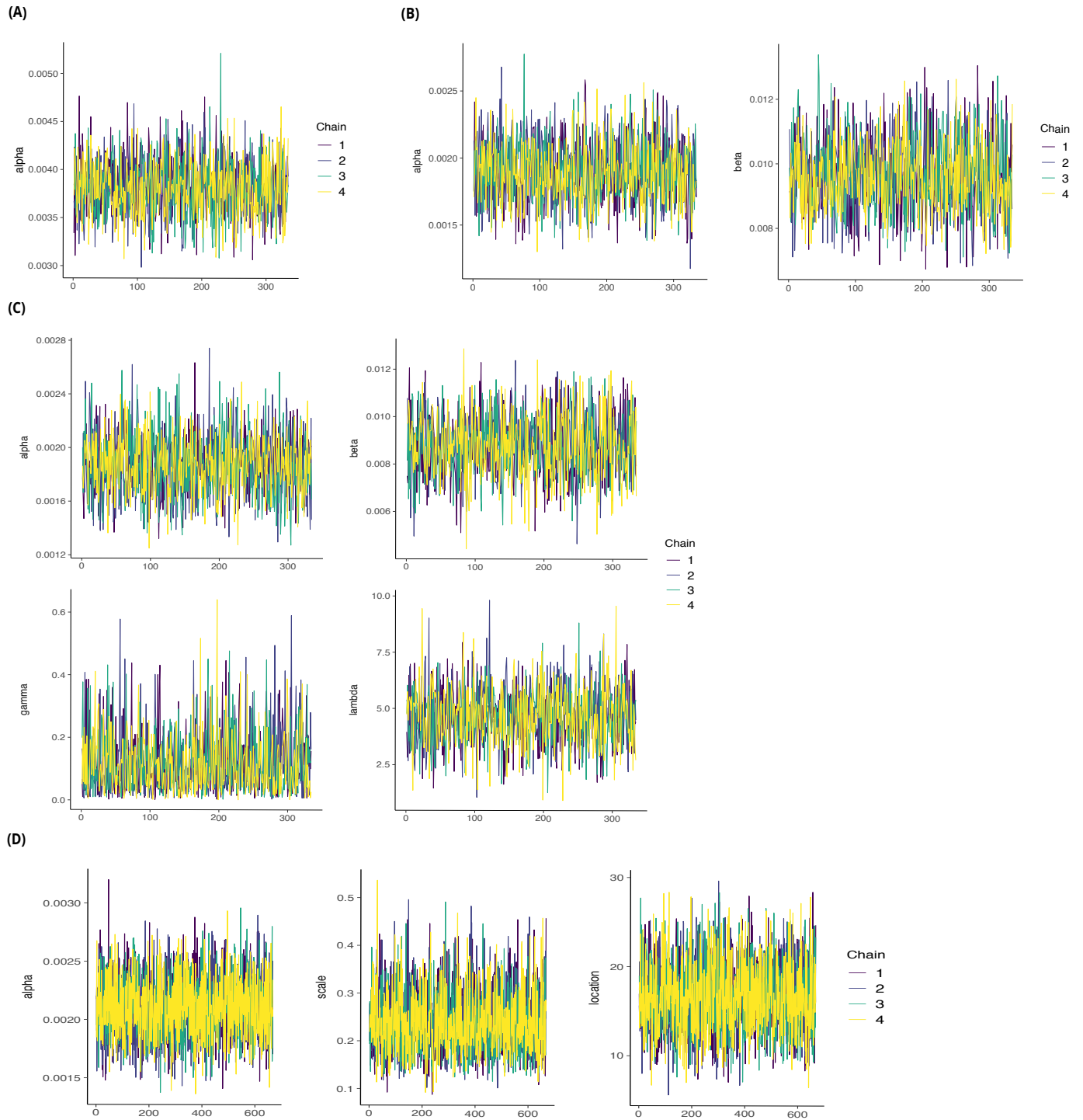
## 858 Figure Supplements



**Figure 2- figure supplement 1.** Posterior Chains from Hamiltonian Markov Chain Monte Carlo fitting using Stan for risk factor model A (see Table 2). The parameters are the intercept  $\alpha$  and 14 slopes for covariates  $\beta$ s. The Gelman-Rubin diagnostic ( $\hat{R}$ ) is  $< 1.01$  for all parameters. The tail effective sample size ( $ESS_{tail}$ ) ranges from 1948-2499 and the bulk effective sample size ( $ESS_{bulk}$ ) ranges from 1808-2350.

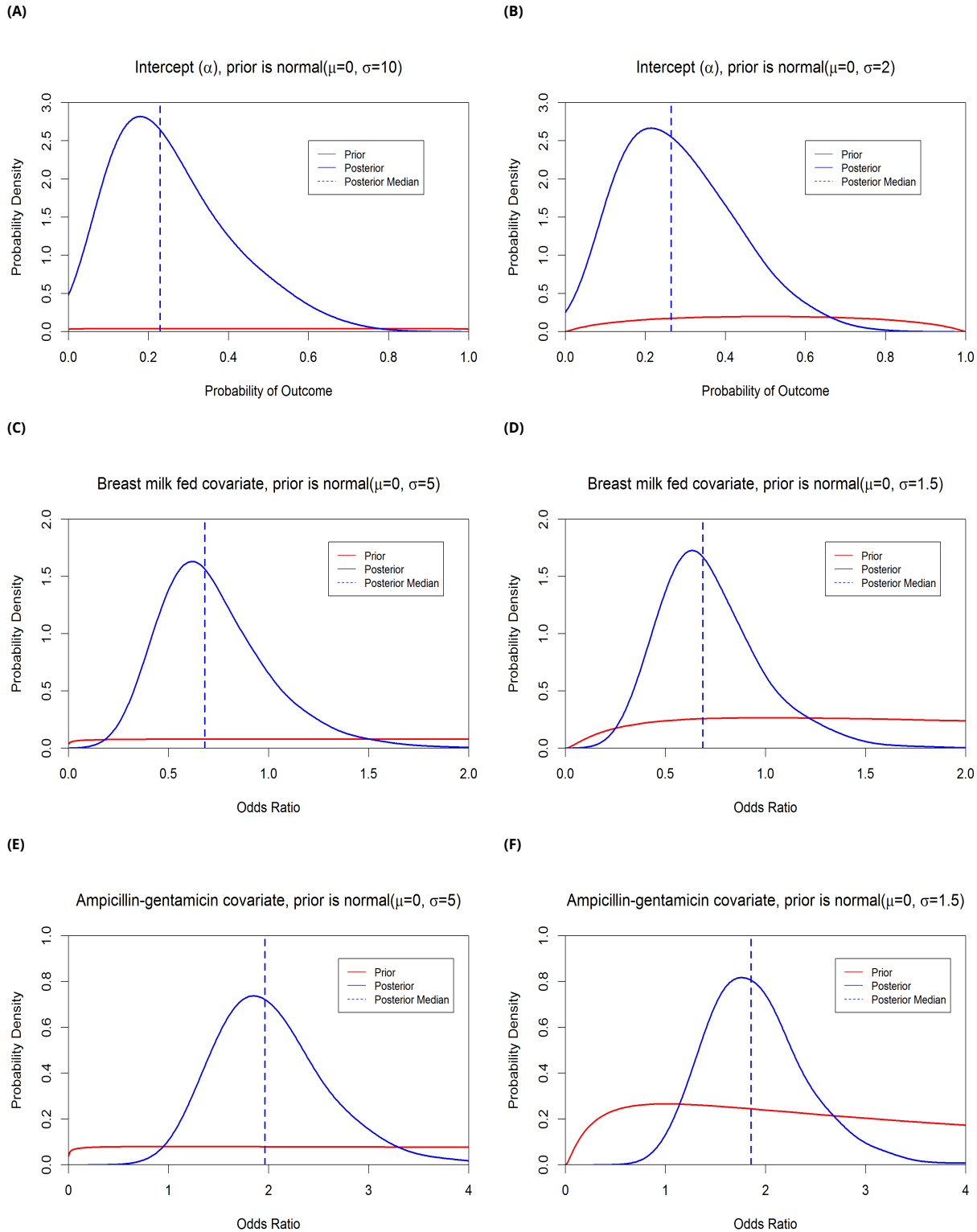


**Figure 2- figure supplement 2.** Estimates from Risk Factor Model 1 with variable priors. The intercept ( $\alpha$ ) prior is varied from normal( $\mu=0, \sigma=10$ ) to normal( $\mu=0, \sigma=2$ ). This changes the median posterior probability from 0.23 to 0.26 (panels A and B). Covariate prior distributions were varied from normal( $\mu=0, \sigma=5$ ) to normal( $\mu=0, \sigma=1.5$ ). The median odds ratio for the effect of breast feeding on the risk of acquisition / detection of resistant *Klebsiella pneumoniae* changes from 0.68 to 0.69 (C-D), and the effect of taking ampicillin + gentamicin within the past 48 hours changes from 2.0 to 1.9 (E-F). Note that priors are fit on the log-odds scale and that the prior and posterior distributions shown in this figure have been logit-transformed (A-B) or exponentiated (C-F).



**Figure 3- figure supplement 1.** Posterior Chains from Hamiltonian Markov Chain Monte Carlo fitting for *Klebsiella* transmission models (see Table 3). A) Transmission model 1; intercept [alpha]  $\hat{R}=1.001$ ,  $ESS_{bulk}=1359$ ,  $ESS_{tail}=1171$ . B) Transmission model 2; alpha and beta;  $\hat{R}$  ranges from 1.001-1.002,  $ESS_{bulk}$  ranges from 1144-1298,  $ESS_{tail}$  ranges from 1247-1282. C) Transmission model 3; alpha, beta, gamma and lambda.  $\hat{R}$  ranges from 1.000-1.003,  $ESS_{bulk}$  ranges from 1274-1396,  $ESS_{tail}$  ranges from 1180-1285. D) Transmission model 4. alpha and beta hyper-parameters scale/shape 1, location/shape 2 and transmission parameters for each ST (62, not shown),  $\hat{R}$  ranges between 0.999-1.003,  $ESS_{bulk}$  ranges from 1687-2887 and  $ESS_{tail}$  ranges from 2105-2796.





**Figure 3- figure supplement 2** Estimation of  $\lambda$  from Transmission Model 3 under different prior assumptions. The more informative prior, half-normal( $\mu=1, \sigma=2$ ), gave a median parameter estimate of 4.7 (panel A) corresponding to an environmental half life of 3.6 hours (panel C). The less informative prior, half-normal( $\mu=0, \sigma=5$ ), gave a median parameter estimate of 6.7 (panel B) corresponding to an environmental half life of 2.5 hours (panel C). The only published estimate of resistant *Klebsiella pneumoniae* half life on surfaces is 12 hours ( $\lambda=1.39$ , see Methods), therefore the informative prior gives more weight to biologically plausible values of  $\lambda$ . The qualitative interpretation of the model (that environmental contamination with *K. pneumoniae* decays quickly) is unchanged with either prior, as is WAIC.