SUPPLEMENTARY TABLES

Table S1: Most common CH types by ESC-R status (only including the first isolate collected per individual).

CH type		ESC-R (n=278)	ESC-S (n=1008)
H30 (40-30)		83 (29.9)	47 (4.7)
Other ST-131-associated types			
	40-41	7 (2.5)	26 (2.6)
	40-22	3 (1.1)	2 (0.2)
	40-27	3 (1.1)	8 (0.8)
Other common types			
	37-27	10 (3.6)	1 (0.1)
	26-5	10 (3.6)	9 (0.9)
	11-54	7 (2.5)	11 (1.1)
	35-27	7 (2.5)	105 (10.4)
	38-41	3 (1.1)	102 (10.1)
	14-27	0 (--)	47 (4.7)
	24-10	0 (--)	46 (4.6)
All others		145 (52.2)	604 (59.9)

Table S2: Raw numbers used for the ST131-*H30* and *H30Rx* prevalence estimates. These numbers can include repeat isolates from a given patient if they were collected within 15 days of the first isolate. ESC-R = extended-spectrum cephalosporin-resistant. ESC-S = extended spectrum cephalosporin-susceptible.

		West	Midwest 1	Midwest 2	East	All sites
ESC-R		116	100	37	81	334
	H30 (all)	29	40	13	37	119
	H30Rx	25	22	10	25	82
ESC-S		348	299	111	240	998
	H30 (all)	13	16	3	15	47
	H30Rx	1	2	0	2	5
Total # of *E. coli*[a]		3021	6668	2163	9947	21799

[a]This count was provided by each study center and was used to calculate weights for the overall estimate of prevalence.

Table S3: Analysis of interaction between age and underlying medical condition on risk of ESC-R *H30* infection vs. infection with other ESC-R *E. coli* types using log-binomial regression models

	Age		
	0-5 years	6-21 years	
	RR (95% CI)*	RR (95% CI)*	**RRs (95% CI)* for age 0-5 within strata of underlying medical condition**
Presence of an underlying medical condition	2.07 (1.22-3.52)	1.0 (ref)	2.04 (1.20-3.45)
No underlying medical condition	2.07 (1.18-3.64)	1.11 (0.52-2.36)	1.92 (0.98-3.75)
RRs (95% CI)* for underlying medical condition within age strata	1.00 (0.67-1.49)	0.87 (0.41-1.87)	

Interaction Contrast Ratio (ICR) (95% CI) = -0.10 (-1.67 – 1.03). When interpreting the ICR, deviation from 0 indicates evidence of interaction on the additive scale (see Supplementary Methods).

*RRs adjusted for study center

Table S4: Sensitivity analysis replacing effect of underlying medical condition with effect of underlying urologic condition on risk of *H30* infection vs. infection with other extended spectrum cephalosporin-sensitive *E. coli* using log-binomial regression models

	ESC-S		
	Total effect RR (95% CI)		Direct effect RR (95% CI)
	Crude	Adjusted	Adjusted
Underlying urologic condition	4.66 (2.68-8.21)*	4.32 (2.48-7.65)[a]*	3.49 (1.93-6.39)[b]*

[a] Additional covariates: age (0-5 or 6-21).
[b] Additional covariates: age (0-5 or 6-21, hospitalization in the past 6 months (yes/no), antibiotics in the last 30 days (yes/no), indwelling device (yes/no).
* Confidence interval does not include 1.

SUPPLEMENTARY FIGURES

Figure S1: Schematic dendrogram of ST131-*H30* population structure as interpreted in this study, with associated phenotypic and genotypic characteristics.[1,2] Solid arrows indicate strong associations identified in existing literature, but are not universal; organisms that have undergone recombination or lost mobile genetic elements may not possess these characteristics. ESBL = extended-spectrum beta-lactamase.

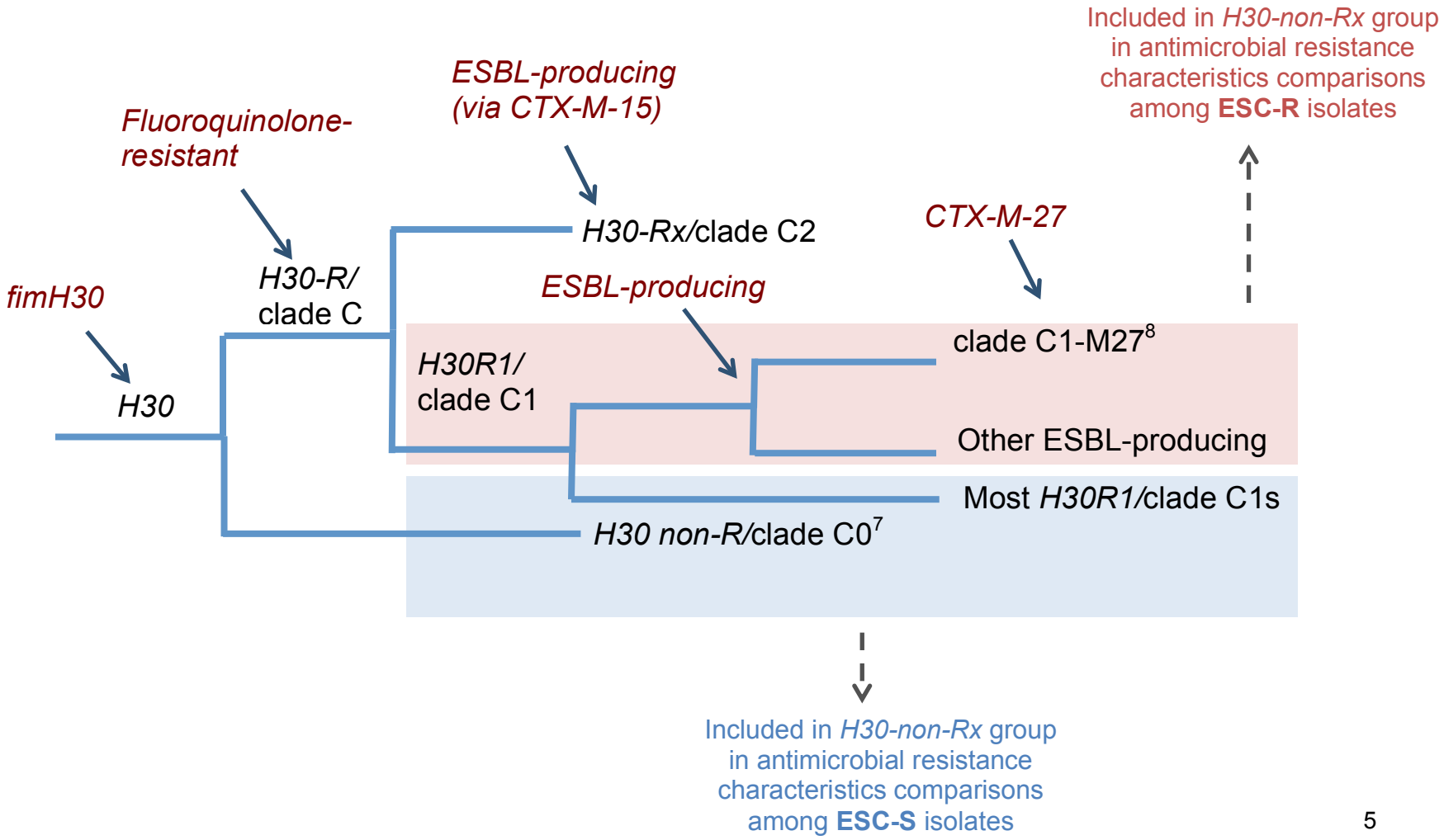


Figure S2: Schematic of subsets of isolates and data used in each presented analyses.

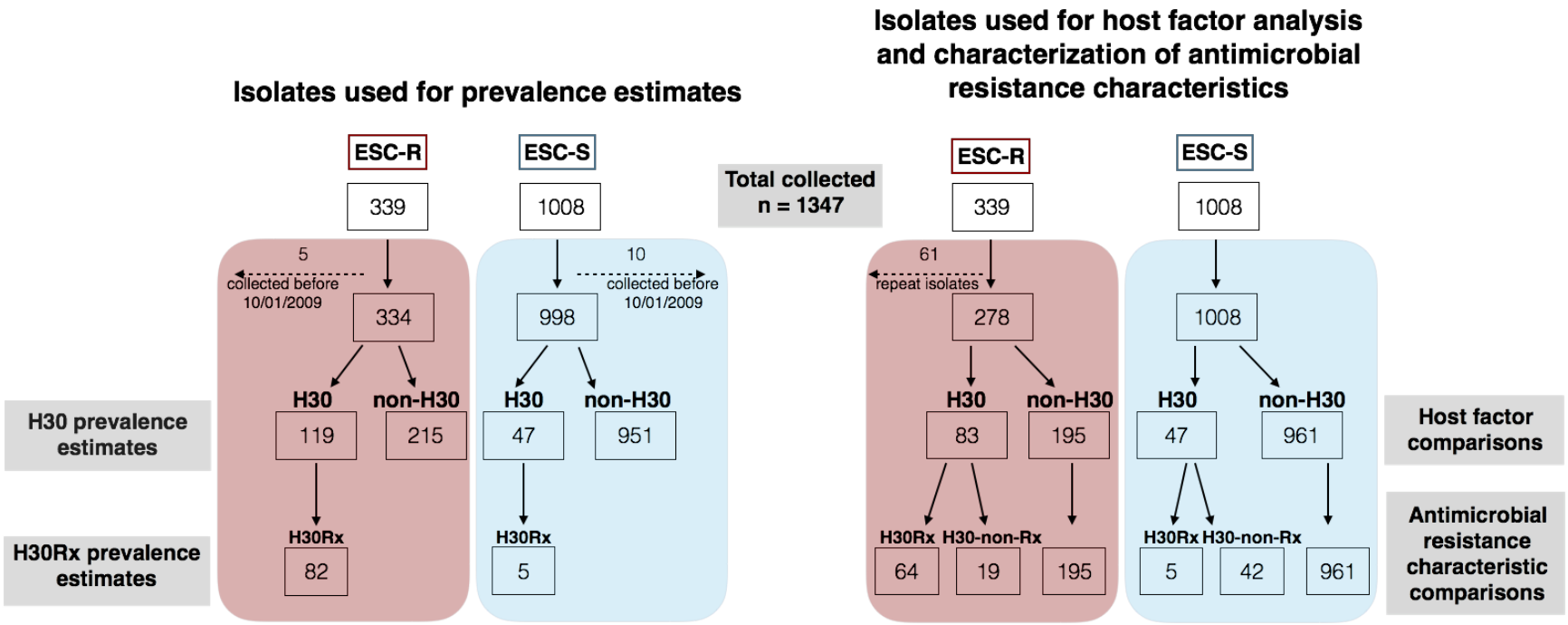
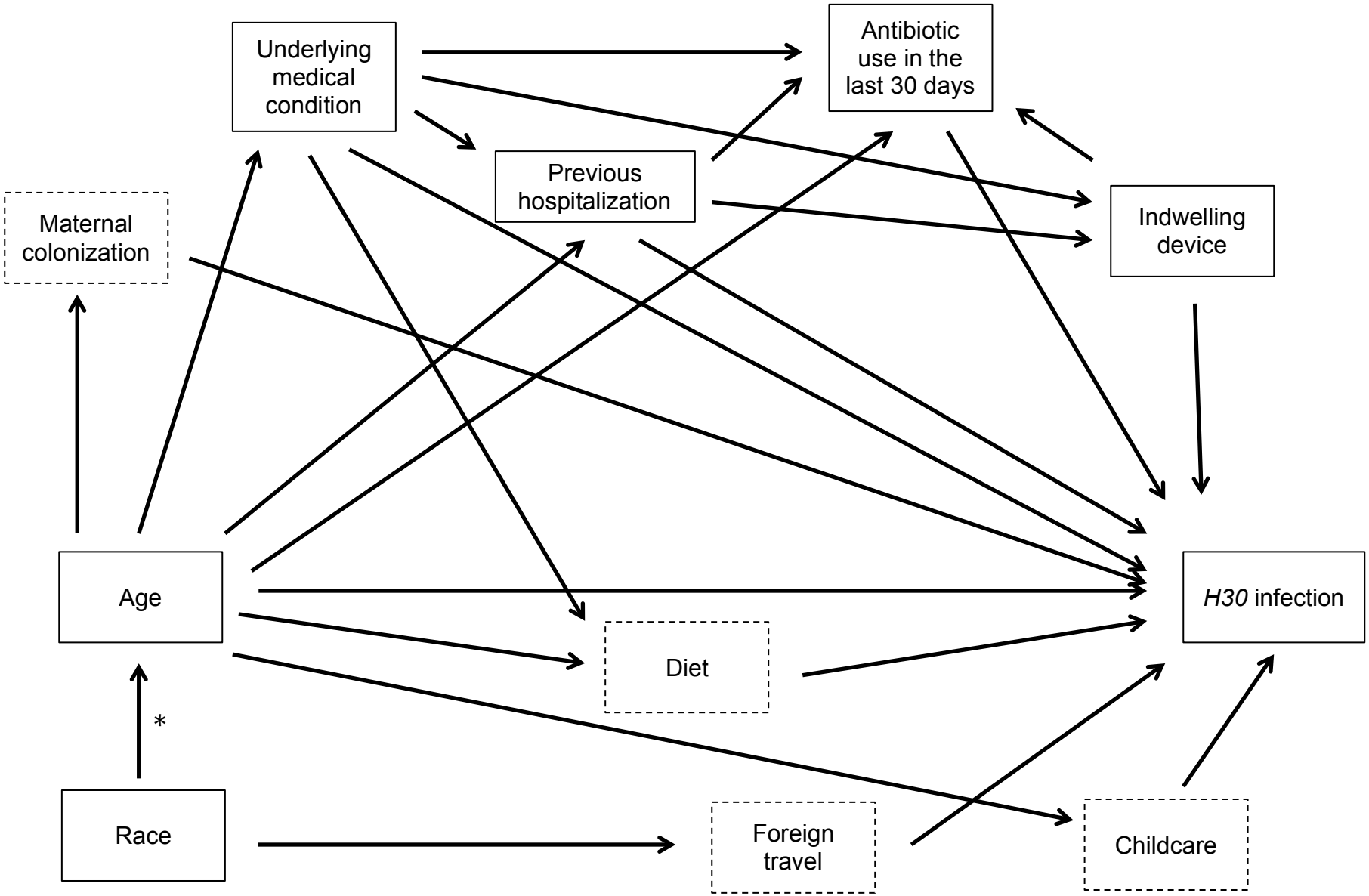


Figure S3: Conceptual framework used for building multivariable models of the relationship between host clinical and demographic factors and ST131-*H30* infection among children with ESC-R extraintestinal *E. coli* infections. Unmeasured variables are surrounded by dashed boxes.



*This association is present in our data, but may not be present in other samples or populations.

Figure S4: Conceptual framework used for building multivariable models of the relationship between host clinical and demographic factors and ST131-*H30* infection among children with ESC-S extraintestinal *E. coli* infections. Unmeasured variables are surrounded by dashed boxes.

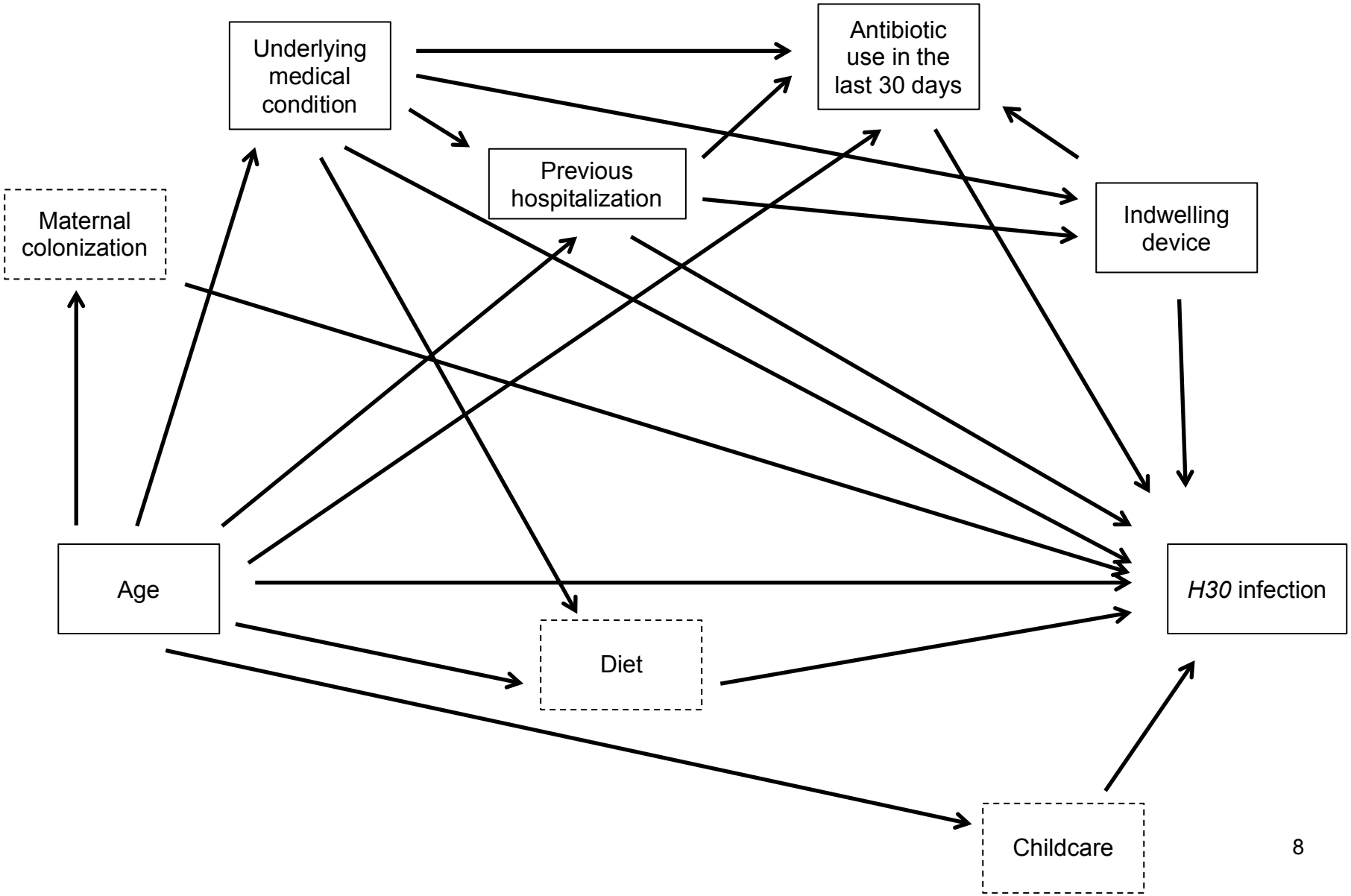
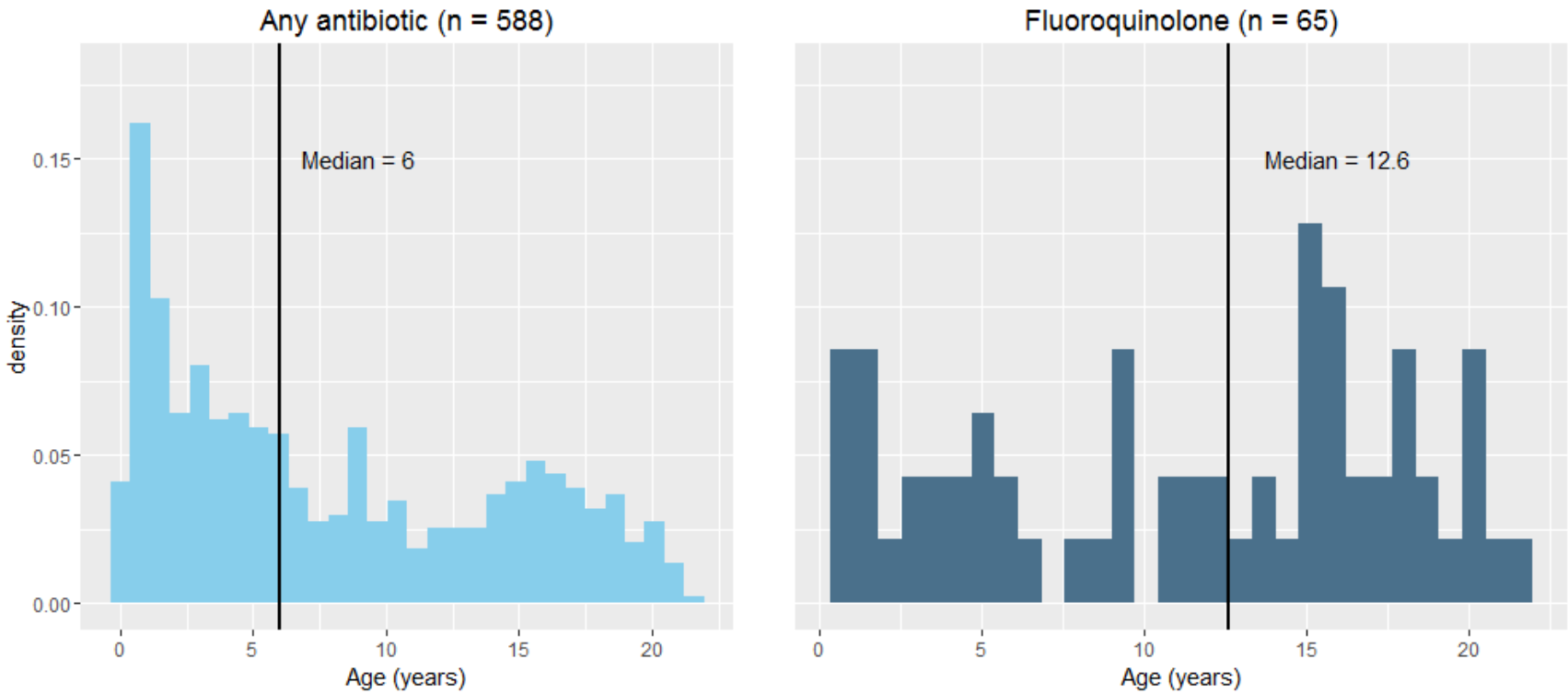


Figure S5: Count of CTX-M-type ESBL determinants and CMY-2 AmpC determinants by study site and by year.*

*Deviation from the vertical line for each year was randomly generated to facilitate visualization, and is not representative of actual temporal variation. Only aggregation by year is permitted by the human subjects approval of this study.

Figure S6: Age distribution of patients that received any antibiotic in the year prior to the collection of the first isolate (left) and of those that received a fluoroquinolone in the year prior to the collection of the first isolate (right) among all patients (N = 1286)



SUPPLEMENTARY METHODS

Statistical methods

Prevalence estimates

The prevalence of *H30* was estimated using the prevalence of *H30* among ESC-R isolates (which were captured completely), the prevalence of *H30* among the collected ESC-S isolates (which were captured partially), and the total number of *E. coli* isolates collected from each clinical microbiology laboratory during the study period, excluding repeat isolates from a given patient if they were collected within 15 days of the first isolate. The 15-day cut point was chosen to reflect the end of a typical course of antibiotic treatment, which is $\leq$ 14 days. This cut-point was prescribed in the original study protocol, and was used by each study site to provide denominator data for the total number of *E. coli* isolates that came through the laboratory during the study period. The total number of *E. coli* isolates from each study hospital was only available between October 1, 2009 and September 30, 2013, so any isolates collected during September 2009 were excluded (Figure S2). After calculating the prevalence of *H30* among ESC-R and ESC-S isolates separately, the overall prevalence among all clinical *E. coli* isolates was estimated using a weighted average of the stratified prevalence estimates, with the weights being the relative proportions of the two mutually exclusive groups (ESC-R and ESC-S) among the total number of *E. coli* isolates reported. All data used for these calculations can be found in Table S2. This approach makes two assumptions: 1) all ESC-R isolates were captured; and 2) the collected ESC-S isolates are representative of all ESC-S isolates.

To estimate uncertainty around these prevalence estimates, we used a resampling-based approach to calculate 95% interval estimates. Since our data were sampled to over-represent ESC-R isolates, in our bootstrap analyses, we weighted each ESC-R observation with the prevalence of ESC-R isolates among all isolates in the population, and each ESC-S observation with the prevalence of ESC-S isolates among all isolates in the population (Table S2). We applied the non-parametric bootstrap procedure using *simpleboot::one,boot* in R with 10,000 bootstrap replicates. We reported the bias-corrected and accelerated bootstrap (BCa) interval for these estimates.[3] The same approach was applied to calculate the prevalence and 95% CI for *H30Rx.*

Host factor analyses

Identification of potential predictors of interest: Variables were first compared using Mantel-Haenszel chi-square tests with stratification by study hospital using the *mantelhaen.test* function in R. When the strata-specific sample sizes were too small for Mantel-Haenszel adjustment or for the chi-square approximation to be valid, Fisher's Exact tests were used via the *fisher.test* function in R. Factors with a p-value of <0.05 were further examined as candidate predictors of interest.

Regression models: The magnitude of the associations between candidate predictors of interest and *H30* infection was examined using univariable and multivariable log-binomial regression models. Log-binomial regression was chosen over logistic regression because the outcome (*H30*) was common among ESC-R isolates; if logistic regression were used, the odds ratio would not approximate the relative risk. The log-binomial regression models were implemented using the *logbin* package in R with the adaptive barrier method selected for maximum likelihood estimation. All multivariable models were adjusted for study hospital. Other covariates were chosen following the conceptual models in Figures S3 and S4 using causal diagram principles.

All covariates were chosen *a priori,* without regard to associations in the data, except for Asian race. In our data, young age was strongly associated with Asian race only among ESC-R isolates (data not shown), and Asian race also displayed an association with *H30* infection among ESC-R isolates. Therefore, we included Asian race (yes/no) as a potential confounder of the association between patient age and *H30* infection among ESC-R isolates. The conceptual framework in Figure S3 highlights the hypothesized mechanism for this observed association.

Interaction analyses: Post-hoc analyses of the mechanistic interaction between categorized patient age and presence of an underlying medical condition were conducted by examining the stratified and joint effects of these two variables. Multivariable log-binomial models adjusting for study hospital were constructed; the adaptive barrier method was selected for maximum likelihood estimation. To examine the joint effects, a 4 level categorical variable representing the combinations of age and underlying medical condition was included as the predictor of interest. We chose to measure the interaction on the additive scale, as this is typically most useful for assessing the public health importance of interactions. Interaction contrast ratios (ICRs) were calculated to measure the extent of interaction on the additive scale: an ICR of 0 represents no interaction, an ICR <0 indicates a negative interaction, and an ICR > 0 indicates a positive interaction.[4,5] Preventative factors were recoded as risk factors, and the stratum with the lowest risk was used as the referent category.[6] Uncertainty in the ICR was estimated via a non-parametric bootstrap approach using the *boot* package in R with antithetic resampling; 1000 replicates were performed and 95% BCa intervals were presented.[3]

Antimicrobial resistance characteristics

Comparisons between *H30* and non-*H30* isolates and *H30Rx* and non-*H30* isolates were quantified using Chi-square tests using the *chisq.test* function in R; Fisher's Exact tests were used when sample size did not allow for a chi-square approximation.

References

1. Banerjee R, Johnson JR. A new clone sweeps clean: The enigmatic emergence of Escherichia coli sequence type 131. *Antimicrob Agents Chemother*. 2014;58(9):4997-5004. doi:10.1128/AAC.02824-14.
2. Mathers AJ, Peirano G, Pitout JDD. The Role of Epidemic Resistance Plasmids and International High-Risk Clones in the Spread of Multidrug-Resistant Enterobacteriaceae. *Clin Microbiol Rev*. 2015;28(3):565-591. doi:10.1128/CMR.00116-14.
3. DiCiccio TJ, Efron B. Bootstrap confidence intervals. *Stat Sci*. 1996;11(3):189-228. doi:10.1214/ss/1032280214.
4. Greenland, S., Lash, T. L., and Rothman, K. J. (2008). “Concepts of interaction,” chapter 5. In: Modern Epidemiology, K. J. Rothman, S. Greenland, and T. L. Lash (Eds.). 3rd Edition. Philadelphia, PA: Lippincott Williams and Wilkins.
5. VanderWeele TJ, Knol MJ. A Tutorial on Interaction. *Epidemiol Method*. 2014;3(1):33-72. doi:10.1515/em-2013-0005.
6. Knol MJ, VanderWeele TJ, Groenwold RHH, Klungel OH, Rovers MM, Grobbee DE. Estimating measures of interaction on an additive scale for preventive exposures. *Eur J Epidemiol*. 2011;26(6):433-438. doi:10.1007/s10654-011-9554-9.
7. Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, et al. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of Escherichia coli ST131. *MBio*. 2016;7(May):1-12. doi:10.1101/039123.
8. Matsumura Y, Pitout JDD, Gomi R, et al. Global Escherichia coli Sequence Type 131 Clade with bla CTX-M-27 Gene. *Emerg Infect Dis*. 2016;22(11):1900-1907. doi:10.3201/eid2211.160519.