



Joint multidrug-resistance information in Indian antimicrobial-resistance surveillance: bounds from marginal and pairwise susceptibility data and a temporal validation

OVERVIEW: Antimicrobial-resistance surveillance usually reports susceptibility one antibiotic at a time, although multidrug resistance is a joint property of each isolate. Using historical Indian ATLAS records for *Klebsiella pneumoniae* and *Escherichia coli*, this study quantified how much information about nonsusceptibility to five antibiotics at once is lost when only separate percentages are kept, and whether drug-pair frequencies and an early-period relationship recover it. The findings support retaining joint information and validating analytical shortcuts across time, but do not support reducing direct susceptibility testing.

AUTHORS: Amritash Patra (Independent researcher; sole author)

KEYWORDS: antimicrobial resistance; multidrug resistance; surveillance; susceptibility testing; joint probability bounds

ABSTRACT: Antimicrobial-resistance surveillance commonly reports susceptibility separately for each antibiotic, yet multidrug resistance is a joint property of the same isolate. This study examined how much information about nonsusceptibility to all five of a set of antibiotics can be recovered from separate antibiotic frequencies, whether drug-pair information narrows the resulting uncertainty, and whether an early-period relationship remained valid later. Historical Indian ATLAS records for *Klebsiella pneumoniae* and *Escherichia coli* were analysed. Of 641 target-species records, 544 had complete results for five prespecified antibiotics. For each organism, the observed all-five frequency was compared with an independence estimate (the product of the five separate frequencies), with the exact feasible range implied by the separate frequencies alone, and with the narrower range obtained after incorporating drug-pair frequencies. For *K. pneumoniae*, the observed all-five frequency was 16.54%, compared with 1.17% under independence; the range from separate frequencies alone was 0–22.06%, and the pair-informed range was 14.34–19.12%. For *E. coli*, the corresponding values were 5.51%, 0.18%, 0–6.99% and 5.15–5.88%. In temporal validation, a later *K. pneumoniae* burden of 25/72 (34.7%) exceeded both the early upper bound of 16.5% and the reported

97.5th bootstrap percentile of 21.5%. The findings support retaining joint information and validating analytical shortcuts across time, but they do not establish current national prevalence, clinical utility, or a basis for reducing direct antimicrobial susceptibility testing.

INTRODUCTION:

Antimicrobial resistance (AMR) is a major public-health concern, and surveillance systems are intended to characterise resistance patterns over time and across settings. India has established national surveillance activities through the Indian Council of Medical Research Antimicrobial Resistance Surveillance and Research Network and the National Centre for Disease Control National AMR Surveillance Network. Contemporary surveillance reports and global AMR reporting emphasise the importance of standardised, interpretable resistance data for detecting patterns and informing public-health action.¹²³

A central statistical issue is that multidrug resistance is a joint state. For an isolate tested against five antibiotics, five separate percentages describe the prevalence of nonsusceptibility to each drug, but they do not specify how those outcomes co-occur within isolates. Two populations can therefore share identical separate



percentages while differing in the frequency of isolates nonsusceptible to all five drugs. The quantity of interest is a higher-order joint probability, whereas the reported inputs may consist only of lower-dimensional summaries.

The ATLAS database is relevant to this question because it stores isolate-level susceptibility information with associated metadata, and analyses that reduce observations to separate resistance frequencies can discard information about co-occurrence. The original ATLAS study describes vectors of antibiotic measurements for individual isolates and documents substantial heterogeneity in the availability of antibiotic testing across records.¹ This creates a natural setting in which to compare what can be learned from complete joint observations with what remains identifiable after information is compressed.

If only separate probabilities are known, the five-way intersection probability is generally not identified. For binary nonsusceptibility indicators X_1 to X_5 with probabilities p_1 to p_5 , the independence approximation is the product of the five probabilities. By contrast, the sharp lower and upper bounds for the five-way intersection given only the separate probabilities are $\max(0, p_1 + p_2 + p_3 + p_4 + p_5 - 4)$ and $\min(p_1, \dots, p_5)$, respectively. Partial knowledge of lower-dimensional joint probabilities can further constrain the feasible set of higher-order joint distributions; these are instances of Fréchet-type bounds and dependence uncertainty.²

Pairwise data are therefore potentially informative without being sufficient in every

situation. With five antibiotics there are ten drug pairs, and retaining their joint frequencies adds constraints on the 32 possible binary susceptibility states. Whether those constraints materially reduce uncertainty is an empirical question, as is whether a relationship estimated in one population remains useful in another. Temporal transportability is distinct from mathematical identifiability: a relationship can be informative in one period and still fail to characterise a later population.

The objective of this study was to quantify the gap between separate-percentage summaries and the observed five-drug joint state in historical Indian ATLAS records for *K. pneumoniae* and *E. coli*, to assess how much that gap could be constrained by pairwise information, and to evaluate whether an early-period relationship remained consistent with later data. A secondary objective was to stress-test a proposed reduced-testing shortcut without treating its success or failure as evidence for a clinical change in susceptibility testing. Because the pairwise frequencies add constraints on the same five-dimensional distribution, the feasible all-five range was expected to be no wider after pairwise information was incorporated; temporal validation was deliberately treated as a stress test rather than assuming that an early-period relationship would remain stable.

RESULTS

Dataset composition

The analysis included 641 target-species records from historical Indian ATLAS data for *K. pneumoniae* and *E. coli*. Of these, 544 records had complete results for the five prespecified antibiotics and formed the complete-case five-drug cohort (Table 1). The source did not provide a species-specific denominator table or a per-antibiotic testing availability table, so these were not reconstructed. See <https://doi.org/10.1038/s41467-022-30635-7> for details.

² Roger B. Nelsen et al., “Best-Possible Bounds on Sets of Bivariate Distribution Functions,” *Journal of Multivariate Analysis* 90, no. 2 (2004): 348–358, <https://doi.org/10.1016/j.jmva.2003.09.002>.



Discordance between independence estimates and observed multidrug states

For *K. pneumoniae*, the observed proportion of isolates in the all-five nonsusceptible state was 16.54%, whereas the independence estimate obtained by multiplying the five separate percentages was 1.17%. For *E. coli*, the corresponding values were 5.51% observed and 0.18% under independence (Figure 1; Table 2). The product-of-marginals shortcut therefore gave substantially smaller values than the observed higher-order state in both organisms, a discrepancy of approximately 14-fold for *K. pneumoniae* and 31-fold for *E. coli* as described in the source analysis.

Feasible joint-state bounds from separate percentages

Using the five separate percentages alone, the exact feasible range was 0–22.06% for the *K. pneumoniae* all-five state and 0–6.99% for *E. coli*. These intervals show that the separate frequencies do not uniquely determine the joint state: many joint distributions can share the same five separate frequencies while assigning different mass to the all-five cell.

Information gained from pairwise frequencies

Adding every drug-pair frequency narrowed the feasible range to 14.34–19.12% for *K. pneumoniae* and 5.15–5.88% for *E. coli* (Figure 2; Table 2). The observed values, 16.54% and 5.51%, respectively, lay within their pair-informed ranges. In the analysed data, pairwise information therefore imposed much tighter constraints on the five-way joint state than separate percentages alone, without requiring a claim that pairwise information is universally sufficient.

Temporal validation

The temporal stress test exposed a lack of transportability of the early-period constraint. For *K. pneumoniae*, the early upper bound was 16.5% and the reported 97.5th bootstrap percentile was 21.5%. In the later 2015–2016 period, the observed

all-five burden was 25/72, or 34.7%, which exceeded both early-period reference quantities (Figure 3; Table 3). This comparison shows that an informative relationship in the early data cannot be assumed to remain valid in a later population without independent validation; it does not establish a universal time trend.

Testing availability and shortcut evaluation

The source noted a sharp change in testing availability across periods. It also reported that a two-drug candidate shortcut failed a pre-specified safety rule because there were too few informative events, and that an adaptive testing strategy saved no tests. Because the underlying event counts and the exact rule definition were not available, no numerical performance estimate is assigned to the shortcut.

Exploratory genomic assessment

No quantitative genomic result was available beyond the statement that a resistance-enriched Indian whole-genome-sequencing and susceptibility-testing cohort was examined as an exploratory check. This component was therefore treated as exploratory and was not used to support a mechanistic conclusion.

DISCUSSION

The central result is a distinction between marginal information and joint information. Five separate drug-specific percentages identify five one-dimensional margins, but the all-five nonsusceptibility state is a five-dimensional joint probability. Unless the dependence structure is specified or additional joint information is retained, the higher-order probability is not identified. The independence product is one value implied by a particular dependence assumption; it is not a consequence of the separate percentages themselves.

The observed discrepancies illustrate why this matters for surveillance measurement. The product of five separate percentages was 1.17% for *K. pneumoniae* against an



observed 16.54%, and 0.18% for *E. coli* against an observed 5.51%. The appropriate interpretation is not that independence is inherently a poor model in all AMR settings, but that its validity cannot be assumed solely from correct separate percentages. A model based on separate percentages can appear numerically precise while remaining highly uncertain about the higher-order state.

The feasible-range analysis provides a complementary perspective. When only separate probabilities are retained, the admissible all-five state spans a broad interval. Pairwise frequencies supply information about dependence structure and consequently reduce the set of joint distributions consistent with the summaries. In the analysed data, the pair-informed intervals contained the observed all-five values and were substantially narrower than the intervals from separate percentages alone. This is evidence that pairwise summaries can preserve information lost under purely marginal reporting, not evidence that pairwise data alone generally identify all higher-order relationships.

The temporal analysis adds a separate caution. The later *K. pneumoniae* burden of 34.7% exceeded both the early upper bound of 16.5% and the reported early-information bootstrap 97.5th percentile of 21.5%, so the constraint that appeared informative in earlier data did not bound the later observation. This should not be reframed as proof of a universal temporal trend: the data are historical and not nationally representative, and the source provides no model that separates biological evolution from changes in testing composition, population structure or ascertainment.

The results also separate four concepts that are easily conflated. Mathematical identifiability asks whether the available summaries uniquely determine the target quantity. Empirical association asks whether variables co-occur in observed data. Prediction asks whether an estimated

relationship can forecast a new sample with adequate calibration and event support. Clinical decision-making asks whether changing testing or treatment on the basis of that prediction has acceptable consequences. This study establishes a mathematical information-loss result and an empirical temporal stress test; it does not establish a clinically validated prediction rule and does not justify replacing direct susceptibility testing.

The study has several strengths. It used isolate-level surveillance records rather than relying only on national summary rates; it compared direct observation with both a simple independence approximation and explicit feasible bounds; it added pairwise information rather than assuming it away; and it included an out-of-period validation test rather than evaluating a shortcut only on the data used to construct it. The failure of the two-drug shortcut is also informative, because a proposed efficiency gain should be rejected when event support is inadequate.

The study also has limitations. It used historical data that are not nationally representative and therefore does not estimate current Indian AMR prevalence. The temporal comparison is a validation stress test rather than a causal analysis, and the later exceedance could reflect several sources of population and measurement change that cannot be separated here. The genomic component is exploratory and quantitatively underspecified, and no genomic mechanism or gene-level result is claimed. Essential reproducibility details were not preserved: the identities of the five antibiotics, breakpoint authority and version, handling of intermediate, susceptible and resistant categories, detailed cohort denominators, pairwise frequency inputs, bootstrap replicate count and resampling unit, random seed, the optimisation implementation and the exact two-drug shortcut definition. The analysis does not establish treatment efficacy, clinical utility, a safe reduction in



susceptibility testing, or a basis for selecting one antibiotic over another.

The main implications concern surveillance data architecture and analytic validation. Where isolate-level susceptibility has already been measured, preserving the joint susceptibility vector, tested or not-tested status, timing, specimen and site context, testing method and breakpoint version can reduce avoidable information loss. Where line lists cannot be shared, richer joint summaries, including governed pairwise frequencies, may preserve more information than separate percentages. Any reduced-testing strategy should be frozen before evaluation and tested on temporally or geographically independent data, with explicit thresholds for event support and uncertainty. These are surveillance-data recommendations, not treatment recommendations. Future work should first resolve the missing reproducibility details, after which the analysis could be re-run independently, subjected to sensitivity analyses across breakpoint systems and cohorts, and evaluated in additional time periods or independent Indian datasets.

In conclusion, separate antibiotic-specific percentages did not uniquely determine the joint five-drug nonsusceptibility state in historical Indian surveillance data. The all-five state differed markedly from the independence product, while pairwise information substantially narrowed the feasible range in both organisms. A relationship that appeared informative in the early period did not constrain the later *K. pneumoniae* observation, underscoring the need for independent temporal validation. The defensible conclusion is a surveillance-measurement one: preserving joint susceptibility information and testing shortcuts on new data are important, but these results do not support reducing clinical susceptibility testing.

MATERIALS AND METHODS

Study design and data source

This was a retrospective, observational computational analysis of historical Indian records in the ATLAS antimicrobial susceptibility database. The source analysis identified 641 target-species records, of which 544 had complete results for five prespecified antibiotics. The target organisms were *K. pneumoniae* and *E. coli*. No complete breakdown of these records by organism, year or antibiotic was available, and none is inferred here. The early period was defined as 2004–2009 and the later period as 2015–2016.

Antibiotic panel and susceptibility classification

Five prespecified antibiotics were used for the five-drug analysis. Their names and identifiers, and the exact susceptibility-classification rule, breakpoint authority and breakpoint version, were not preserved in the source; no antibiotic identity or breakpoint system is assigned or inferred here, and the reported classifications were not recast under a different standard.

Missingness and testing availability

Testing was not complete for every target-species record. Naturally missing results and the change in testing availability across periods were examined, but per-drug missingness counts and a formal missing-data model were not available, and no imputation was described. The five-drug analysis therefore used complete cases only.

Representation of joint resistance

For each isolate and antibiotic j , X_j was defined as a binary indicator of the study-defined nonsusceptible state. The five-drug joint state of interest was $q_{11111} = P(X_1=1, X_2=1, X_3=1, X_4=1, X_5=1)$, and the separate frequencies were $p_j = P(X_j=1)$. The observed q_{11111} was calculated from isolates with all five results available.



Independence approximation and feasible bounds

The independence approximation was defined as $q_{ind} = p_1 \times p_2 \times p_3 \times p_4 \times p_5$, a modelling shortcut that assumes mutual independence of the five indicators. The marginal-only feasible set consisted of all non-negative joint distributions over the 32 possible five-drug binary states whose one-dimensional margins equalled the reported p_j ; q_{11111} was bounded by minimising and maximising its value over this set. For five events, the sharp bounds are $\max(0, \sum p_j - 4)$ and $\min(p_j)$. The numerical intervals reported here were taken directly from the source analysis.

Pair-informed constraints

Every drug-pair frequency (ten pairs for five antibiotics) was incorporated as an additional constraint. The pair-informed feasible set contained joint distributions satisfying the five one-dimensional margins together with the reported pairwise probabilities, non-negativity and unit total probability, and the minimum and maximum q_{11111} over this set gave the pair-informed interval. The pairwise input table and exact computational implementation were not retained, so only the mathematical formulation and resulting ranges are reported.

Temporal validation

Temporal validation focused on the *K. pneumoniae* all-five nonsusceptibility burden. The later period contained 25/72 isolates in the all-five state (34.7%). The early upper bound (16.5%) and the early-information bootstrap 97.5th percentile (21.5%) were compared descriptively with this value to assess whether the early-period constraint transported to the later population. The bootstrap replicate count, resampling unit, random seed and any formal decision threshold were not preserved, so the 21.5% value is reported as a source-derived percentile rather than a new confidence interval or hypothesis test.

Reduced-testing shortcut evaluation

A two-drug candidate shortcut was evaluated against a pre-specified safety rule, together with an adaptive-testing strategy. The identities of the two drugs and the formal definitions of the rule and algorithm were not retained and are not inferred here.

Exploratory genomic assessment

A resistance-enriched Indian whole-genome-sequencing and susceptibility-testing cohort was referred to in the source as an exploratory check. No cohort size, accession identifiers, genomic variables or quantitative results were available, so no gene-level, mechanistic or predictive genomic claim is made.

Statistical reporting

The primary quantities were descriptive probabilities, feasible bounds and a temporal exceedance comparison. No inferential p-values, confidence intervals, regression coefficients or causal effects are reported. Feasible ranges represent mathematical feasibility under the stated information, not statistical uncertainty. All numerical values were preserved from the source analysis.

ACKNOWLEDGMENTS

Funding: No funding was received for this work.

Data availability: The analysis used historical Indian records from the ATLAS antimicrobial susceptibility database, described in the original ATLAS publication as a large raw MIC database with associated patient metadata.³ The underlying records are not redistributed here, and no extraction file or derived dataset was retained; only the cohort definition and reported results are reproduced.

³ Pablo Catalán et al., "Seeking Patterns of Antibiotic Resistance in ATLAS, an Open, Raw MIC Database with Patient Metadata," *Nature Communications* 13, no. 1 (2022): 2917, <https://doi.org/10.1038/s41467-022-30635-7>.



Code availability: Analysis code was used to produce the reported results. No public repository is currently available, and no release identifier, commit, software-version record or public archive was retained.

Ethics: No ethics approval was obtained. The work is a secondary analysis of historical ATLAS records and did not recruit participants, collect new specimens or conduct an intervention.

Conflict of interest: The author had not yet confirmed a conflict-of-interest declaration at the time this manuscript was prepared.

Author contributions: Amritash Patra: conceptualisation, data selection and curation, methodology, formal analysis, software, validation, interpretation and writing of the original manuscript. The research question, data selection, analytical design, mathematical bounds, quantitative results, interpretation and scientific decisions were developed and verified by the author.

Supplementary material: No supplementary material accompanies this version.

REFERENCES

1. World Health Organization. "Global Antibiotic Resistance Surveillance Report 2025: WHO Global Antimicrobial Resistance and Use Surveillance System (GLASS)." Geneva: World Health Organization, 2025. <https://iris.who.int/handle/10665/383106>.

2. Indian Council of Medical Research. "Annual Report 2024: ICMR Antimicrobial Resistance Research and Surveillance Network (AMRSN)." 8th ed. New Delhi: Indian Council of Medical Research, 2025. https://www.icmr.gov.in/icmrobject/uploads/Report/1763981012_icmramrsnannualreport2024.pdf.

3. National Centre for Disease Control. "National Antimicrobial Resistance Surveillance Network (NARS-Net) Annual Report 2024: Reporting Period January–December 2024." New Delhi: Ministry of Health and Family Welfare, Government of India, 2025. https://ncdc.mohfw.gov.in/wp-content/uploads/2025/09/Final_Annual-Report-2025_Jan-to-Dec-2024.pdf.

4. Catalán, Pablo, et al. "Seeking Patterns of Antibiotic Resistance in ATLAS, an Open, Raw MIC Database with Patient Metadata." *Nature Communications* 13, no. 1 (2022): 2917. <https://doi.org/10.1038/s41467-022-30635-7>.

5. Nelsen, Roger B., et al. "Best-Possible Bounds on Sets of Bivariate Distribution Functions." *Journal of Multivariate Analysis* 90, no. 2 (2004): 348–358. <https://doi.org/10.1016/j.jmva.2003.09.002>.

6. Gheewalla, Nazneen, et al. "Genomic Landscape of Antimicrobial Resistance in India: Findings from a Multi-Species Surveillance Study." *npj Antimicrobials and Resistance* 4 (2026): 13. <https://doi.org/10.1038/s44259-026-00185-9>.

Figures and Figure Titles/Captions



Figure 1. Marginal reporting can obscure joint multidrug resistance

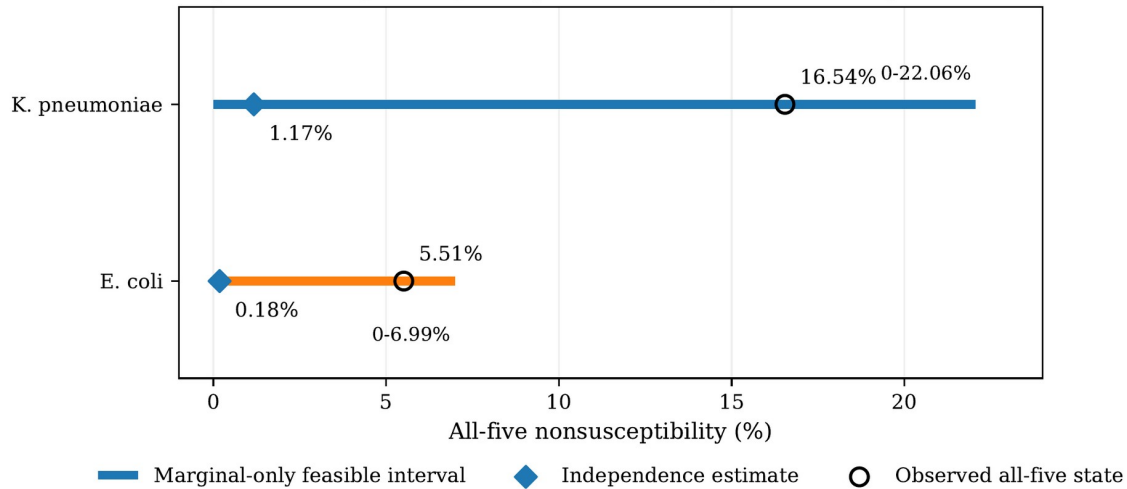


Figure 1. Marginal reporting can obscure joint multidrug resistance.

Horizontal intervals show the exact feasible ranges for the five-drug all-nonsusceptible state implied by the five separate percentages alone. Diamonds show the independence estimate (product of the five separate percentages) and open circles show the observed all-five state. *K. pneumoniae*: separate-percentage range 0–22.06%, independence 1.17%, observed 16.54%. *E. coli*: range 0–6.99%, independence 0.18%, observed 5.51%. The intervals represent mathematical feasibility, not confidence intervals; values were taken from the source analysis (n = 544 complete five-drug records across both species).

Figure 2. Pairwise information constrains the feasible all-five state

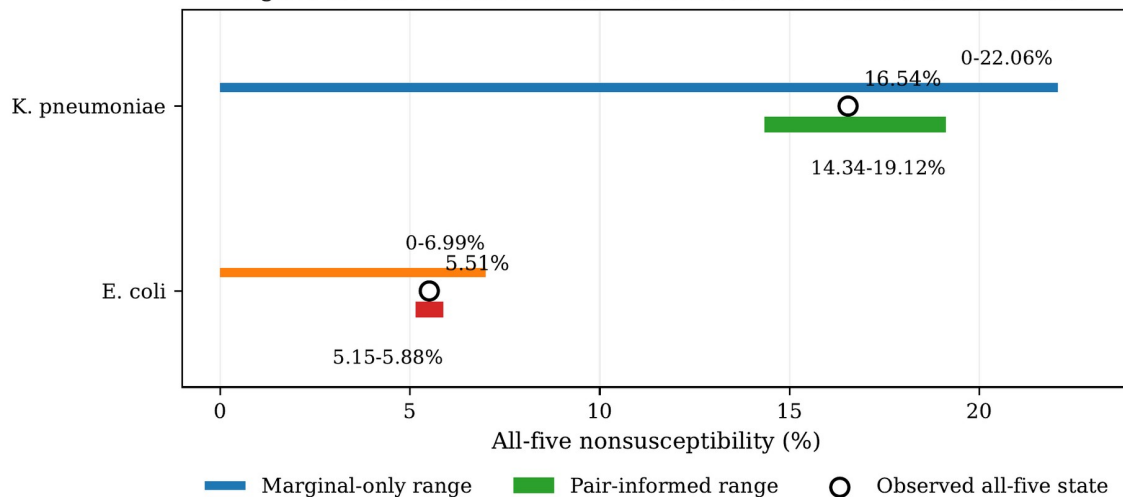


Figure 2. Pairwise information constrains the feasible all-five state.

Horizontal intervals compare the feasible ranges from separate percentages alone with the narrower ranges obtained after incorporating every drug-pair frequency. *K. pneumoniae*: 0–22.06% versus 14.34–19.12%, observed 16.54%. *E. coli*: 0–6.99% versus 5.15–5.88%, observed 5.51%. The intervals are feasible joint-state ranges rather than statistical confidence intervals, and the figure does not imply that pairwise information is sufficient for arbitrary higher-order distributions.

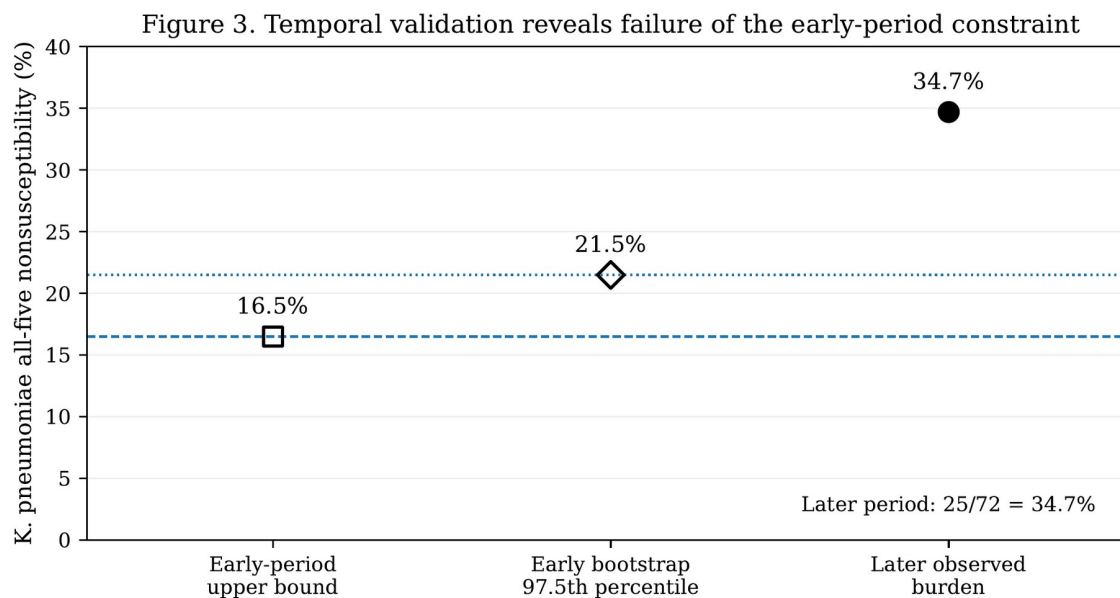


Figure 3. Temporal validation reveals failure of the early-period constraint.

The early-period upper bound (16.5%) and the reported early-information bootstrap 97.5th percentile (21.5%) are compared with the later *K. pneumoniae* all-five nonsusceptibility burden of 25/72 (34.7%). The later value lies above both early reference quantities. This is a temporal transportability stress test, not proof of a universal resistance trend or a causal temporal effect; bootstrap replicate count and resampling unit were not preserved.

Tables with Titles/Captions

Study component	Source-supported information	Reproducibility status
Target organisms	<i>Klebsiella pneumoniae</i> ; <i>Escherichia coli</i>	Supported by source
Target-species records	641	Supported by source
Complete five-antibiotic records	544	Supported by source
Early period	2004–2009	Supported by source
Later period	2015–2016	Supported by source
Five antibiotics	Five prespecified antibiotics; exact identities not preserved	Source limitation
Susceptibility classification	Exact rule and breakpoint authority/version not preserved	Source limitation
Bootstrap specification	97.5th percentile reported; replicate count, resampling unit and seed not preserved	Source limitation
Code repository	No public repository available; analysis code was used to generate the results	Source limitation

Table 1. Study dataset and analytic cohorts.

Records were drawn from historical Indian ATLAS data. “Source limitation” marks details that were not preserved in the source analysis.

Organism	Observed all-five state	Independence estimate	Range from separate percentages only	Pair-informed range
<i>K. pneumoniae</i>	16.54%	1.17%	0–22.06%	14.34–19.12%



Organism	Observed all-five state	Independence estimate	Range from separate percentages only	Pair-informed range
<i>E. coli</i>	5.51%	0.18%	0–6.99%	5.15–5.88%

Table 2. Observed and information-constrained all-five nonsusceptibility estimates.

The independence estimate is the product of the five separate nonsusceptibility frequencies. Ranges are mathematical feasibility ranges, not confidence intervals. All values were taken from the source analysis.

Validation component	Source-reported result	Interpretation boundary
Early <i>K. pneumoniae</i> upper bound	16.5%	Reference constraint from earlier information
Early bootstrap 97.5th percentile	21.5%	Reported percentile; bootstrap implementation details missing
Later <i>K. pneumoniae</i> burden	$25/72 = 34.7\%$	Later observation exceeded both early reference quantities
Two-drug shortcut	Failed pre-specified safety rule	Too few informative events; rule details missing
Adaptive testing	Saved no tests	Algorithm details missing
Clinical AST conclusion	Reduction not supported	No clinical recommendation made

Table 3. Temporal validation and shortcut evaluation.

AST, antimicrobial susceptibility testing.