Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Phenotypic and Molecular Approaches to Antimicrobial Susceptibility Testing in Clinical Laboratories

Fatima Abdulwahab

abdulwahabfatima43@gmail.com

Bridgewater State University

Research Article

Keywords: Keywords: antimicrobial susceptibility testing; molecular diagnostics; phenotypic testing; whole genome sequencing; diagnostic agreement; clinical microbiology

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11850089/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Phenotypic and Molecular Approaches to Antimicrobial Susceptibility Testing in Clinical Laboratories

A reproducible reanalysis of published paired genomic and phenotypic results for Escherichia coli and Klebsiella pneumoniae bloodstream isolates

Fatima Abdulwahab

Independent Researcher, Dallas, Texas, USA

ORCID: 0009-0008-2298-8386 · Correspondence: abdulwahabfatima43@gmail.com

Abstract

Phenotypic antimicrobial susceptibility testing and molecular resistance detection provide related evidence, but their agreement depends on the organism, antimicrobial, comparator and interpretation rules. This study combines a critical methodological synthesis with a reproducible secondary analysis of published paired genomic and phenotypic counts from 143 bloodstream isolates. Fourteen organism–drug strata covered 74 *Escherichia coli* and 69 *Klebsiella pneumoniae* isolates tested against seven antimicrobials. Across 1,001 isolate–drug comparisons, binary agreement was 96.70%. Twelve genomic-susceptible/comparator-resistant results represented 1.20% of all comparisons but 3.39% of the 354 comparator-resistant results. Twenty-one genomic-resistant/comparator-susceptible results represented 2.10% of all comparisons and 3.25% of the 647 comparator-susceptible results. Organism–drug patterns were uneven. *E. coli*–ceftazidime accounted for eleven of the twenty-one genomic-resistant/comparator-susceptible results. The only comparator-resistant *K. pneumoniae*–meropenem result was predicted susceptible, producing a highly uncertain stratum estimate. An analytical prevalence scenario showed how the probability of comparator resistance after a genomic-susceptible result changes even when sensitivity and specificity are held constant. These findings demonstrate why global agreement should accompany class-specific denominators, uncertainty intervals and a clear account of discrepancy resolution. The historical dataset supports a methodological demonstration rather than an assessment of present-day systems. Clinical implementation requires current breakpoints, validated workflows, appropriate quality control and evidence that results reach treatment decisions in time.

Keywords: antimicrobial susceptibility testing; molecular diagnostics; phenotypic testing; whole-genome sequencing; diagnostic agreement; clinical microbiology

1 Introduction

Antimicrobial susceptibility testing links laboratory measurements with a treatment decision. A phenotypic test observes growth or inhibition under specified conditions, while a molecular method identifies a resistance determinant or a sequence pattern associated with a response. Both can contribute to patient care, but a clinically useful result depends on more than whether two reports have the same category. It also depends on the organism, specimen, antimicrobial, testing conditions, breakpoint version and whether the result becomes available when it can change treatment [20, 27, 28].

The distinction is easiest to see when the methods disagree. A genomic prediction of susceptibility may reflect the absence of a known determinant in a database. Resistance can still arise through an unrecognised mechanism, a missing sequence region or an expression effect outside the prediction rule. A detected determinant can also coexist with a susceptible phenotype under a particular test condition or interpretive threshold. Phenotypic results themselves can vary because of inoculum, medium, incubation, reading uncertainty or an organism–drug limitation of the platform [13, 20, 23]. A discrepancy therefore requires investigation before assigning its cause.

Summary accuracy can conceal that investigation. If most tested isolates are susceptible, a high overall agreement can coexist with a clinically important fraction of resistant isolates receiving a susceptible result. Recent guidance on categorical agreement likewise stresses how category definitions affect comparisons [26]. The denominator changes the interpretation of the same discrepancy count. Reporting the count per all comparisons describes its frequency in the evaluated collection. Reporting it among comparator-resistant isolates describes a conditional miss rate. Both can be informative, provided the label and denominator are explicit.

This study addresses two questions. How should phenotypic and molecular methods be compared within a clinical laboratory workflow? How do organism–drug strata, denominator choice and uncertainty change the interpretation of published paired results? A historical dataset was selected because it reports extractable contingency counts, describes its discrepancy-resolution process and includes both concordant and discordant organism–drug strata. The new contribution is the reconstruction and analysis of those counts, including conditional discordance and prevalence scenarios. It is not an independent evaluation of the original sequencing algorithm or a validation of current clinical platforms. A companion reanalysis applies a similar denominator-focused approach to resistance-marker data from a commercial blood-culture panel (Abdulwahab, manuscript in preparation); the present article concerns genome-based prediction against phenotypic testing.

2 Phenotypic and molecular evidence

2.1 Phenotypic susceptibility testing

Broth microdilution measures growth across a defined concentration series and is an important reference approach for bacterial susceptibility testing. Disk diffusion reports inhibition-zone diameters interpreted using method-specific criteria. Gradient diffusion estimates an inhibitory concentration from growth around a strip, while automated platforms use defined panels and instrument-specific algorithms. Each approach requires suitable inoculum preparation, organism identification and quality-control procedures [11, 20]. Broader reviews of antimicrobial activity methods help explain the distinction between research inhibition experiments and validated diagnostic susceptibility procedures [3].

A minimum inhibitory concentration is a laboratory measurement under specified conditions. Clinical categories incorporate additional considerations, including antimicrobial exposure and the infection context. Under current EUCAST terminology, S indicates susceptibility at a standard dosing regimen, I indicates susceptibility with increased exposure and R indicates resistance

despite increased exposure [15]. The I category has represented susceptible, increased exposure since the 2019 change. A historical study that combined intermediate with resistant cannot be relabelled as a modern EUCAST comparison without the underlying measurements and the relevant species–drug criteria.

Rapid phenotypic approaches shorten the time needed to observe a response. EUCAST RAST work established procedures and reading-time-specific interpretation for selected organism–drug combinations directly from positive blood cultures [1, 19]. Other systems use imaging, microfluidics or related readouts [28]. Their value depends on how often they provide an interpretable result, the time from specimen collection to reporting, and the laboratory's ability to act on a discordant or indeterminate result. A short instrument run cannot by itself establish a short clinical turnaround time.

2.2 Molecular detection and genomic prediction

Targeted amplification can detect a defined gene or mutation rapidly. Its scope is constrained by the targets included and the validation linking those targets to a particular organism–drug interpretation. Genome sequencing interrogates a broader set of determinants, but depends on sequence quality, assembly or read mapping, allele interpretation and database curation. ResFinder, PointFinder, AMRFinderPlus and CARD exemplify different resources and approaches to resistance detection and interpretation [2, 7, 16, 17, 31]. A detected marker should retain its identity and the rule connecting it to an antimicrobial prediction.

Studies in *S. aureus* and other species demonstrate that genomic prediction performance is specific to the organisms and agents evaluated [9, 18]. The earlier EUCAST review and the updated report describe opportunities as well as unresolved issues in translating genomic information into clinical susceptibility predictions [13, 23]. A clinical comparison across bacterial isolates also illustrates the importance of testing genotype–phenotype concordance in the intended setting [22]. A result based on a validated high-penetrance determinant can have a different evidentiary basis from a prediction that depends on several interacting loci. Database content and software changes can alter a result even when the input sequence is unchanged.

Sharing paired data requires identifiers that link the same isolate across phenotype, sequence and metadata, as well as test method, measured result, interpretation and version [10]. Absence of a marker should be distinguished from failure to obtain adequate sequence information. Likewise, a phenotypic result outside a method's validated scope should remain visible as a limitation. Table 1 summarises complementary uses and limitations across the major approaches.

Table 1. Phenotypic and molecular approaches in a laboratory workflow

Approach	Primary information	Interpretive limitation
Broth microdilution	Growth across defined concentrations	Requires standardised conditions and appropriate interpretation
Disk or gradient diffusion	Inhibition zone or estimated inhibitory concentration	Agent and organism limitations, reading uncertainty

Approach	Primary information	Interpretive limitation
Automated phenotypic system	Panel-based susceptibility results	Platform-specific scope, rules and verification
Targeted molecular assay	Presence of selected determinants	Unincluded or unrecognised mechanisms can be missed
Genome-based prediction	Broader determinant and sequence information	Database, algorithm, quality and organism–drug validation

2.3 From result generation to clinical impact

Analytical performance is one part of clinical utility. A result can be correct yet arrive after a patient's treatment has already been optimised. It can also arrive quickly but require a stewardship response or a clinician familiar with the organism and drug. Evidence from molecular rapid diagnostic testing and rapid phenotypic trials indicates that workflow and implementation influence treatment effects [4, 5, 25]. Most recently, the multinational FAST trial found that rapid phenotypic testing directly from positive blood cultures shortened the time to antibiotic escalation or de-escalation but was not superior to standard testing for the day-30 ranked clinical outcome [6]. The present study does not estimate patient benefit or turnaround time. It uses published counts to examine how analytical comparisons should be communicated before an implementation claim is made.

3 Methods

3.1 Design and data provenance

This study is a secondary methodological analysis of aggregate contingency counts, accompanied by a focused literature synthesis. It is not a systematic review or meta-analysis. The empirical source is the openly accessible study by Stoesser and colleagues, published in 2013 [24]. The source compared genomic predictions with a phenotypic comparison standard in 74 *E. coli* and 69 *K. pneumoniae* bloodstream isolates. The collection included resistant isolates and susceptible controls and was not a consecutive contemporary clinical sample. Its 143 isolates generated seven antimicrobial comparisons each.

The source used duplicate BD Phoenix testing and additional gradient diffusion to investigate disagreements, including discordance between duplicate phenotypic runs and between genomic prediction and concordant phenotypic results. In unflagged cases, the concordant automated phenotype contributed the comparison category. The resulting comparison standard therefore combined methods and depended partly on which cases triggered further investigation. It is described here as a comparator, with no assumption that it established biological truth in every case.

The full-text XML was downloaded from Europe PMC (PMC3772739) on 19 November 2026. Tables 3 and 4 of the source supplied four counts per organism–drug pair. These were extracted by a Python script using table identifiers DKT180TB3 and DKT180TB4 (Appendix A). All rates were recalculated from the four cell counts, independently of the summary sensitivity and specificity

printed in the source tables. Counts were checked against the published species totals and against 74 or 69 observations per drug. The *K. pneumoniae* ceftriaxone and ceftazidime rows are identical (42, 1, 2 and 24); this was confirmed against the source, whose discrepancy listing attributes the discordant results for both drugs to the same isolates. All fourteen strata were retained in the primary reconstruction. No genomes were reanalysed, no new organisms were cultured and no patient-level records were obtained.

3.2 Metrics and denominator definitions

With genomic resistance as the positive prediction and comparator resistance as the positive outcome, the four cells are true positive, false positive, false negative and true negative relative to that comparator. The terms are computational labels rather than independent adjudication of the cause of a mismatch. Binary agreement equals the sum of the matching cells divided by all comparisons. Sensitivity equals genomic-resistant/comparator-resistant results divided by all comparator-resistant results. Specificity equals genomic-susceptible/comparator-susceptible results divided by all comparator-susceptible results.

The two discordance directions were reported using both total and conditional denominators. Genomic S/comparator R divided by all comparisons is a total-comparison frequency. The same numerator divided by comparator-resistant results is a conditional susceptible-discordance rate. Genomic R/comparator S has the corresponding total-comparison and comparator-susceptible rates. Table 2 gives the explicit definitions. Current CLSI terminology distinguishes discrepancy between systems from error against a reference method and defines the resistant denominator for a very major error rate [12]. Because the source comparator combines automated and selected gradient-diffusion results, the manuscript uses directional discordance as its primary language.

Table 2. Metrics and their denominators

Measure	Numerator	Denominator
Binary agreement	Genomic S/comparator S plus genomic R/comparator R	All comparisons
Sensitivity for comparator R	Genomic R/comparator R	All comparator R
Specificity for comparator S	Genomic S/comparator S	All comparator S
Susceptible discordance, overall	Genomic S/comparator R	All comparisons
Susceptible discordance, conditional	Genomic S/comparator R	All comparator R
Resistant discordance, overall	Genomic R/comparator S	All comparisons
Resistant discordance, conditional	Genomic R/comparator S	All comparator S

Species-level and combined rates were calculated by summing counts. These summaries weight strata by their number of comparisons. They are not separate independent samples of 518 or 483 patients. The same isolate contributes several drugs, and aggregate tables do not reveal within-isolate dependence. Pooled estimates are descriptive. No pooled independence-based confidence interval or formal test of a species difference was used. Drug-specific intervals for the susceptible-discordance rate used the Wilson score method, with the comparator-resistant count as

denominator [21, 30]. These intervals quantify binomial uncertainty within a stratum under independent-isolate assumptions.

3.3 Sensitivity and analytical scenarios

K. pneumoniae–amoxicillin was included to reproduce the entire reported count set. However, *K. pneumoniae* is intrinsically resistant to amoxicillin, and its source row is a measurement-level comparison that does not incorporate that clinical interpretation. A sensitivity analysis therefore excluded the stratum from the pooled summary, and both versions are reported with equal prominence. The resulting change illustrates how a biologically or clinically inappropriate comparator stratum can influence a global metric. It provides no current prescribing interpretation for amoxicillin in *K. pneumoniae*.

The second analysis applies Bayes' rule to the species-level sensitivity and specificity. For assumed resistance prevalence p, the probability of comparator resistance after a genomic-susceptible result is $(1 - \text{sensitivity}) \times p$ divided by $[(1 - \text{sensitivity}) \times p + \text{specificity} \times (1 - p)]$. The prevalence varies from 1% to 80% while sensitivity and specificity are held constant. This is an analytical scenario, not a calibrated clinical risk model. The assumptions can fail when different populations have different resistance mechanisms, specimen characteristics or testing quality.

Reporting decisions were informed by STARD and the domains of diagnostic-accuracy bias considered in QUADAS-2 [8, 29]. The study does not claim a formal systematic assessment of the source study. The complete reproduction script is given in Appendix A. The analysis used Python, lxml, NumPy, pandas and Matplotlib.

4 Results

4.1 Reconstruction and binary agreement

All fourteen strata were reconstructed, yielding 1,001 isolate–drug comparisons. There were 354 comparator-resistant and 647 comparator-susceptible results. Matching categories occurred in 968 comparisons, giving 96.70% agreement. The corresponding species-level figures were 97.10% for *E. coli* and 96.27% for *K. pneumoniae*. Table 3 lists the extracted counts so that the denominators can be verified directly.

Table 3. Reconstructed historical counts by organism and antimicrobial

Organism	Antimicrobial	G-S/C-S	G-R/C-S	G-S/C-R	G-R/C-R
E. coli	Amoxicillin	23	1	0	50
E. coli	Co-amoxiclav	46	0	0	28
E. coli	Gentamicin	60	0	0	14
E. coli	Ciprofloxacin	48	0	0	26
E. coli	Ceftriaxone	43	1	1	29
E. coli	Ceftazidime	43	11	1	19
E. coli	Meropenem	74	0	0	0
K. pneumoniae	Amoxicillin	0	3	0	66

Organism	Antimicrobial	G-S/C-S	G-R/C-S	G-S/C-R	G-R/C-R
K. pneumoniae	Co-amoxiclav	47	1	0	21
K. pneumoniae	Gentamicin	45	0	1	23
K. pneumoniae	Ciprofloxacin	45	2	4	18
K. pneumoniae	Ceftriaxone	42	1	2	24
K. pneumoniae	Ceftazidime	42	1	2	24
K. pneumoniae	Meropenem	68	0	1	0

G, genomic prediction; C, phenotypic comparator; S, susceptible; R, resistant. Source: Stoesser et al. (2013), Tables 3 and 4.

Figure 1 shows that agreement differed across organism–drug combinations. *E. coli*–ceftazidime had twelve mismatches in 74 comparisons, giving 83.78% agreement. Several strata had no mismatches. Such a result can reflect a successful comparison in the observed collection, but its evidentiary scope depends on the categories represented. For example, the *E. coli*–meropenem stratum contained 74 comparator-susceptible results and no comparator-resistant results. Its perfect agreement evaluates observed matching of susceptible categories and supplies no sensitivity estimate for resistant isolates.

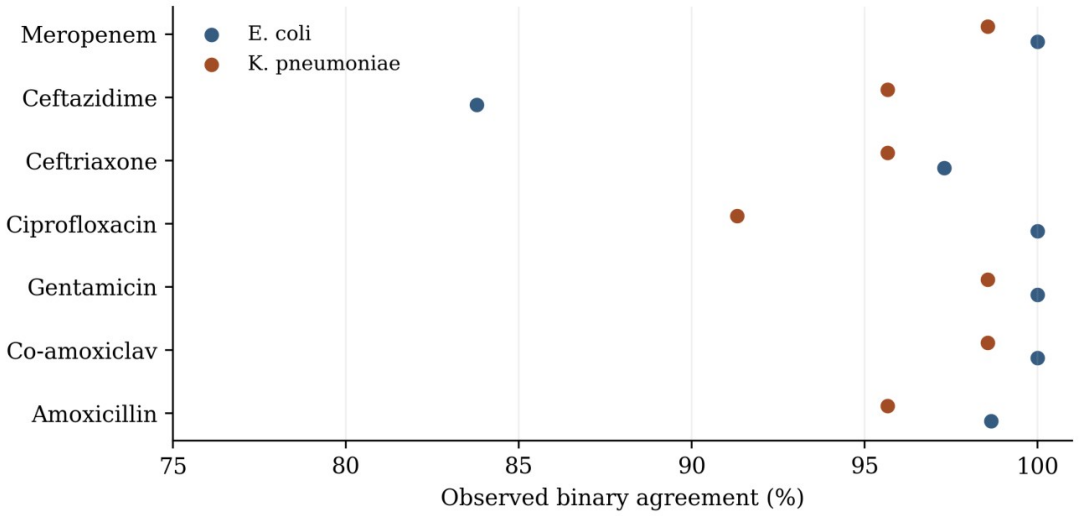


Figure 1. Binary agreement by organism–drug stratum, recalculated from published counts. *E. coli* includes 74 isolates and *K. pneumoniae* 69. Historical categories and comparator procedures are retained.

4.2 The effect of denominator choice

There were twelve genomic S/comparator R results. Dividing by all 1,001 comparisons gives 1.20%, whereas dividing by the 354 comparator-resistant results gives 3.39%. For *E. coli* the corresponding rates were 0.39% and 1.19%. For *K. pneumoniae* they were 2.07% and 5.38%. These are different descriptions of the same numerators. The conditional rate identifies how often the genomic prediction missed the comparator-resistant class in the observed collection.

There were twenty-one genomic R/comparator S results. The all-comparison rate was 2.10% and the comparator-susceptible rate was 3.25%. For *E. coli*, thirteen such results gave rates of 2.51%

among all comparisons and 3.71% among comparator-susceptible comparisons. For *K. pneumoniae*, eight results gave 1.66% and 2.69%. Figure 2 displays all four rates by species and in the combined collection. High agreement and low all-comparison frequencies therefore need the class-specific results to make their interpretation complete.

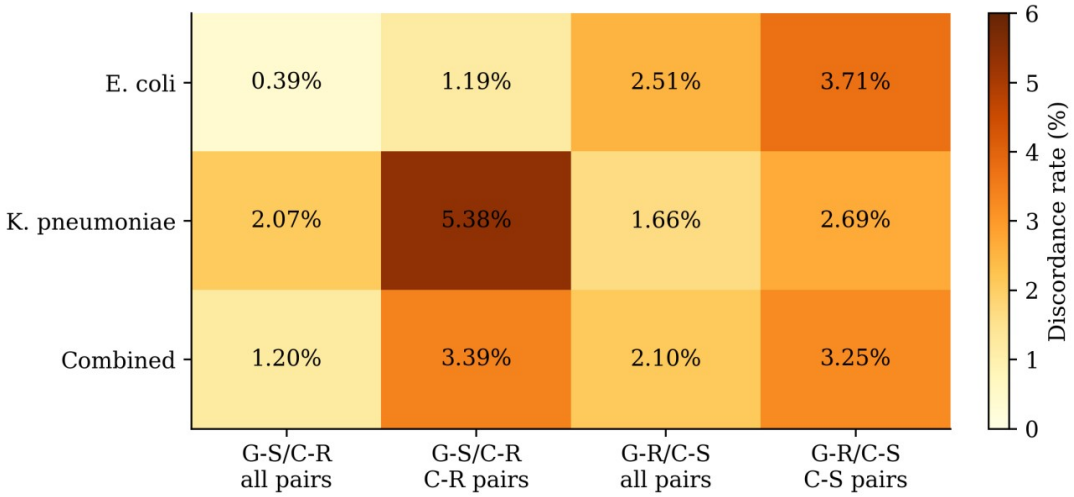


Figure 2. Directional discordance under overall and conditional denominators. G denotes genomic prediction and C the phenotypic comparator. The same numerator can yield different rates depending on the question.

4.3 Stratum size and clinically relevant gaps

Figure 3 reports genomic S/comparator R rates with Wilson intervals within drug strata. *K. pneumoniae*–ciprofloxacin had four such results among 22 comparator-resistant results, a conditional rate of 18.18% (95% CI 7.3–38.5%), the widest clinically relevant gap in the collection. *E. coli*–ceftazidime had one among twenty, or 5.00% (95% CI 0.9–23.6%). *K. pneumoniae*–meropenem had one among one. Its point estimate is 100% (95% CI 20.7–100%), but a single resistant observation gives little precision about performance in a broader resistant population. *E. coli*–meropenem is absent from this conditional-rate plot because its denominator is zero. A zero denominator is distinct from a measured zero rate.

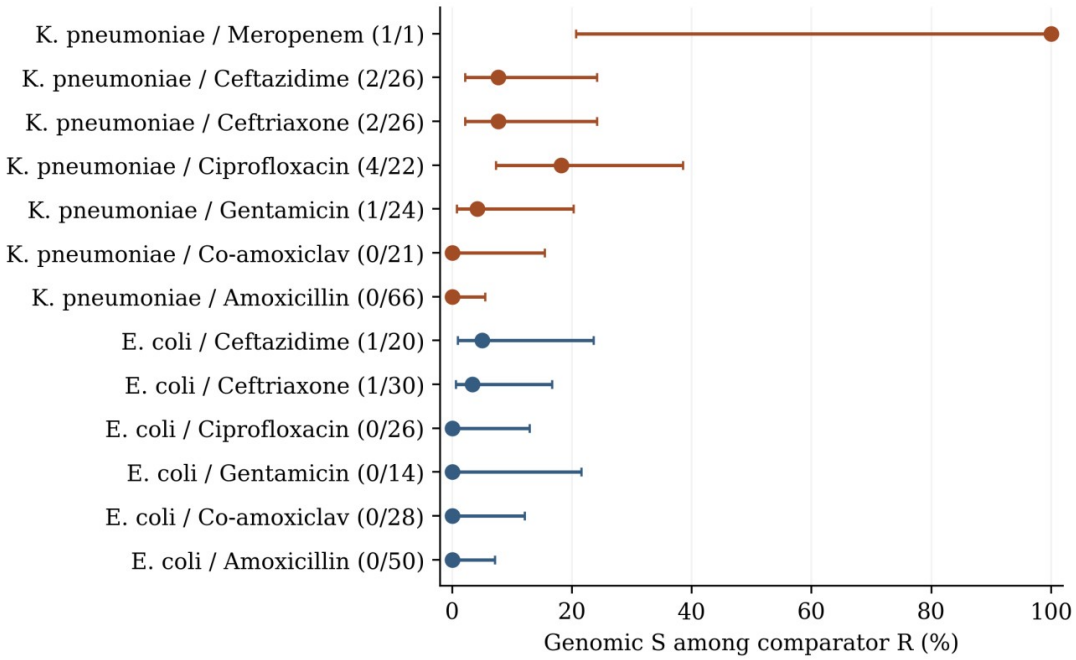


Figure 3. Genomic-susceptible/comparator-resistant results among comparator-resistant isolates. Labels show the numerator and denominator. Whiskers are pointwise 95% Wilson intervals. The *E. coli*–meropenem stratum is excluded because no comparator-resistant isolate was present. This is a stratum summary, not a meta-analysis.

E. coli–ceftazidime contributed eleven of the twenty-one genomic R/comparator S results across the entire collection, or 52.38%. This concentration would be obscured by the pooled specificity of 96.75%. In a laboratory validation report, that pattern warrants a drug-specific review of mechanisms, measurement conditions and the comparator. The aggregate data cannot determine which explanation accounts for an individual discrepant result.

Excluding *K. pneumoniae*–amoxicillin leaves 932 comparisons. Binary agreement changes little, from 96.70% to 96.78%. However, genomic S/comparator R among comparator-resistant results rises from 3.39% to 4.17%, while genomic R/comparator S among comparator-susceptible results decreases from 3.25% to 2.80%. The near-stability of agreement therefore coexists with changes in the directional rates. Table 4 connects these findings to validation and reporting decisions.

Table 4. Analytical findings and implications for validation

Observed feature	Reanalysis result	Reporting or validation implication
Global agreement	968/1,001 = 96.70%	Include the discordance directions and their denominators
Concentrated resistant discordance	*E. coli*–ceftazidime, 11/21 overall G-R/C-S results	Investigate specific organism–drug combinations
Sparse resistant representation	*K. pneumoniae*–meropenem, 1 comparator-R result	Show the denominator and interval; broaden resistant coverage
No resistant representation	*E. coli*–meropenem, 0 comparator-R results	Sensitivity is not estimable

Observed feature	Reanalysis result	Reporting or validation implication
Intrinsic-resistance sensitivity	Exclusion changes conditional G-S/C-R from 3.39% to 4.17%	Check the clinical meaning of included strata

4.4 Prevalence scenarios

Using the historical species-level sensitivity and specificity as fixed inputs, the analytical probability of comparator resistance after a genomic-susceptible result is approximately 0.07% at 5% prevalence for *E. coli* and 1.22% at 50% prevalence. For *K. pneumoniae*, the corresponding probabilities are 0.29% and 5.24%. Figure 4 shows the continuous relationship. These curves illustrate the mathematical dependence on prevalence. Their numerical values cannot be transferred directly to a present clinical population because the input performance comes from a selected historical collection and is pooled across drugs.

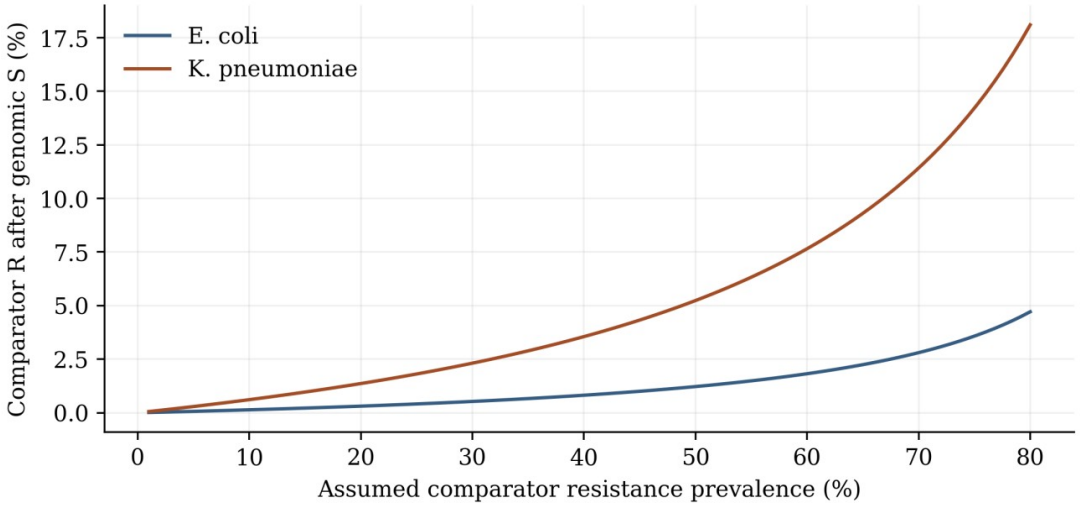


Figure 4. Analytical probability of comparator resistance following genomic susceptibility across assumed prevalence values. Species-level sensitivity and specificity are held fixed at the historical pooled estimates. The curves are scenarios and have no direct clinical calibration.

5 Discussion

5.1 What combined testing can contribute

Phenotypic and molecular testing can answer complementary questions. Phenotyping describes the observed response under a defined assay. Molecular testing can identify a mechanism, support early targeted interpretation and help investigate discordance. An integrated laboratory report benefits from preserving both pieces of evidence rather than reducing them immediately to a single unqualified category. A discrepancy can then be investigated through organism confirmation, repeat measurement, suitable reference testing, sequence quality review and inspection of the determinant-to-drug rule.

The reanalysis demonstrates why that process should be evaluated at the organism–drug level. A pooled agreement near 97% conceals strata with limited resistant representation and one stratum

with concentrated genomic-resistant/comparator-susceptible discordance. The distinction is relevant for verification panels, which need a spectrum of susceptible and resistant isolates that reflects the intended claim. A panel with many easy susceptible results can provide precise evidence about that category while leaving a clinically important resistance mechanism nearly untested.

The original discrepancy-resolution design also limits interpretation. Further testing was triggered by observed disagreement, so the final comparator is partially conditional on initial test results. That design is useful for investigating a feasibility study, but it can miss cases where both methods agree for the same wrong reason. A stronger prospective comparison would apply a prespecified reference procedure to all evaluable isolates, blind interpretation where feasible, and preserve unresolved or indeterminate findings. Discordance adjudication should retain the original results and explain any changes instead of replacing the audit trail.

5.2 Current standards and implementation

As of the literature and standards check on 19 November 2026, EUCAST lists clinical breakpoint tables version 16.1, and CLSI M100 is in its 36th edition with published corrections [11, 14]. The analysis retained the source categories from 2013 because the aggregate tables do not contain the full underlying measurements needed for contemporary reinterpretation. A current evaluation would record the test method, interpretive authority, version, date and relevant notes for every result. A rule that merges EUCAST I with R would erase the present meaning of susceptible, increased exposure and needs explicit justification for a particular research endpoint.

Molecular reporting needs equivalent version control. The same sequence can yield a different inferred phenotype after a database update or a change in allele interpretation. Clinical laboratories should document the software, database release, coverage and identity thresholds, quality requirements and which drug interpretations are validated. The latest EUCAST WGS review emphasises representative populations, reference-quality phenotypes, harmonised interpretation and external quality assessment [23]. These principles are especially relevant when a result is presented as susceptibility rather than simply as detection of a determinant.

Cost and turnaround time require local measurement. A laboratory with limited sequencing capacity may obtain greater practical value from a well-supported phenotypic workflow plus targeted molecular tests than from an expensive broad assay with delayed reporting. Another setting may benefit from sequencing that already supports surveillance and outbreak investigation. A decision should account for staff availability, repeat-test rates, consumables, infrastructure and the pathway from result to treatment review. The current count analysis measures none of these outcomes. Its role is to improve the analytical evidence that accompanies such implementation decisions.

6 Limitations

The empirical analysis draws on one selected historical collection and reuses aggregate data published by another team. It introduces no new isolate testing, sequencing or patient outcomes. Its contribution is methodological re-expression of the counts, with explicit denominators, uncertainty

and sensitivity scenarios. Performance estimates are tied to the source algorithm, resistance distribution and historical interpretation. They cannot rank current commercial systems or justify replacing a validated local procedure.

Within-isolate relationships across drugs are unavailable, so pooled counts describe comparisons rather than independent observations. Formal pooled intervals and comparative significance tests were avoided for that reason. The comparator combines automated phenotypes and selected gradient-diffusion results, and residual discordance is not proof that a genomic method alone was wrong. Intrinsic resistance and expert rules further constrain the interpretation of certain rows. The *K. pneumoniae*–amoxicillin exclusion illustrates one consequence without correcting the dataset retrospectively.

The prevalence scenario assumes fixed performance across populations and drugs. In practice, predictive performance can change with resistance mechanisms and their frequency. Its curves are educational calculations based on stated inputs. A clinical estimate would require an applicable organism–drug model validated in the intended population. Broader claims about the superiority of phenotypic or molecular testing require more representative studies with current methods, complete categories and clinically relevant outcomes.

7 Conclusion

Phenotypic and molecular approaches should be evaluated with attention to their distinct measurements, comparator quality and intended clinical use. Reanalysis of 1,001 historical comparisons showed how high overall agreement can coexist with concentrated drug-specific discordance and limited evidence for a resistant category. Conditional denominators and stratum-specific uncertainty make those gaps visible. An integrated clinical laboratory workflow should preserve mechanism information, validated phenotype results and a traceable route for resolving disagreements.

Declarations

Data and code availability. All source data are publicly available in Stoesser et al. (2013) [24] (Europe PMC full text, PMC3772739). The complete reproduction script is reproduced in Appendix A, and it is also available at [GitHub repository URL] and archived at [Zenodo DOI].

Ethics statement. This study used only published, aggregate, de-identified data. No human participants were recruited and no individual-level records were accessed; ethical approval and informed consent were therefore not required.

Funding. [Confirm: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.]

Competing interests. [Confirm: The author declares no competing interests.]

Author contributions. F.A. conceived the study, extracted and analysed the data, prepared the figures and wrote and revised the manuscript.

Use of AI tools. [State here whether and how AI tools were used, if the target journal requires it.]

References

[1] Åkerlund, A., Jonasson, E., Matuschek, E., Serrander, L., Sundqvist, M., Kahlmeter, G., et al. (2020). EUCAST rapid antimicrobial susceptibility testing (RAST) in blood cultures: validation in 55 European laboratories. *Journal of Antimicrobial Chemotherapy*, 75(11), 3230–3238. https://doi.org/10.1093/jac/dkaa333

[2] Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., et al. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research*, 51(D1), D690–D699. https://doi.org/10.1093/nar/gkac920

[3] Balouiri, M., Sadiki, M., & Ibnsouda, S. K. (2016). Methods for in vitro evaluating antimicrobial activity: a review. *Journal of Pharmaceutical Analysis*, 6(2), 71–79. https://doi.org/10.1016/j.jpha.2015.11.005

[4] Banerjee, R., Giri, A., Komarow, L., Souli, M., Doernberg, S. B., Patel, R., et al. (2025). Impact of rapid antibiotic susceptibility testing for Gram-negative bacteremia varies by pathogen type and resistance: a secondary analysis of the RAPIDS-GN trial. *Microbiology Spectrum*, 13(2), e0178924. https://doi.org/10.1128/spectrum.01789-24

[5] Banerjee, R., Komarow, L., Virk, A., Rajapakse, N., Schuetz, A. N., Dylla, B., et al. (2021). Randomized trial evaluating clinical impact of RAPid IDentification and Susceptibility testing for Gram-negative bacteremia: RAPIDS-GN. *Clinical Infectious Diseases*, 73(1), e39–e46. https://doi.org/10.1093/cid/ciaa528

[6] Banerjee, R., Komarow, L., Li, Y., Mau, D., et al. (2026). Fast antimicrobial susceptibility testing for Gram-negative bacteremia: the FAST randomized clinical trial. *JAMA*, 335(20), 1762–1773. https://doi.org/10.1001/jama.2026.5487

[7] Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., et al. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *Journal of Antimicrobial Chemotherapy*, 75(12), 3491–3500. https://doi.org/10.1093/jac/dkaa345

[8] Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., et al. (2015). STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*, 351, h5527. https://doi.org/10.1136/bmj.h5527

[9] Bradley, P., Gordon, N. C., Walker, T. M., Dunn, L., Heys, S., Huang, B., et al. (2015). Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nature Communications*, 6, 10063. https://doi.org/10.1038/ncomms10063

[10] Chindelevitch, L., van Dongen, M., Graz, H., Pedrotta, A., Suresh, A., Uplekar, S., et al. (2023). Ten simple rules for the sharing of bacterial genotype–phenotype data on antimicrobial resistance. *PLoS Computational Biology*, 19(6), e1011129. https://doi.org/10.1371/journal.pcbi.1011129

[11] Clinical and Laboratory Standards Institute. (2026). *Performance standards for antimicrobial susceptibility testing* (36th ed.; CLSI supplement M100). https://clsi.org/shop/standards/m100/

[12] Clinical and Laboratory Standards Institute. (n.d.). Harmonized terminology database: very major error; overall very major error rate. Retrieved 19 November 2026, from https://htd.clsi.org/listterms.asp

[13] Ellington, M. J., Ekelund, O., Aarestrup, F. M., Canton, R., Doumith, M., Giske, C., et al. (2017). The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clinical Microbiology and Infection*, 23(1), 2–22. https://doi.org/10.1016/j.cmi.2016.11.012

[14] European Committee on Antimicrobial Susceptibility Testing. (2026). *Clinical breakpoint tables, version 16.1*. https://www.eucast.org/bacteria/clinical-breakpoints-and-interpretation/clinical-breakpoint-tables/

[15] European Committee on Antimicrobial Susceptibility Testing. (n.d.). *Definitions of S, I and R*. Retrieved 19 November 2026, from https://www.eucast.org/bacteria/clinical-breakpoints-and-interpretation/definition-of-s-i-and-r/

[16] Feldgarden, M., Brover, V., Fedorov, B., Haft, D. H., Prasad, A. B., & Klimke, W. (2022). Curation of the AMRFinderPlus databases: applications, functionality and impact. *Microbial Genomics*, 8(6), mgen000832. https://doi.org/10.1099/mgen.0.000832

[17] Feldgarden, M., Brover, V., Gonzalez-Escalona, N., Frye, J. G., Haendiges, J., Haft, D. H., et al. (2021). AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Scientific Reports*, 11(1), 12728. https://doi.org/10.1038/s41598-021-91456-0

[18] Gordon, N. C., Price, J. R., Cole, K., Everitt, R., Morgan, M., Finney, J., et al. (2014). Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *Journal of Clinical Microbiology*, 52(4), 1182–1191. https://doi.org/10.1128/JCM.03117-13

[19] Jonasson, E., Matuschek, E., & Kahlmeter, G. (2020). The EUCAST rapid disc diffusion method for antimicrobial susceptibility testing directly from positive blood culture bottles. *Journal of Antimicrobial Chemotherapy*, 75(4), 968–978. https://doi.org/10.1093/jac/dkz548

[20] Jorgensen, J. H., & Ferraro, M. J. (2009). Antimicrobial susceptibility testing: a review of general principles and contemporary practices. *Clinical Infectious Diseases*, 49(11), 1749–1755. https://doi.org/10.1086/647952

[21] Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine*, 17(8), 857–872. https://doi.org/10.1002/(sici)1097-0258(19980430)17:8<857::aid-sim777>3.0.co;2-e

[22] Rebelo, A. R., Bortolaia, V., Leekitcharoenphon, P., Hansen, D. S., Nielsen, H. L., Ellermann-Eriksen, S., et al. (2022). One day in Denmark: comparison of phenotypic and genotypic antimicrobial susceptibility testing in bacterial isolates from clinical settings. *Frontiers in Microbiology*, 13, 804627. https://doi.org/10.3389/fmicb.2022.804627

[23] Samuelsen, Ø., López-Causapé, C., Aarestrup, F. M., Bortolaia, V., Brouwer, M. S. M., Cantón, R., et al. (2026). The role of whole genome sequencing in antimicrobial susceptibility prediction of bacteria: 2025 update from the European Committee on Antimicrobial Susceptibility Testing Subcommittee. *Clinical Microbiology and Infection*, 32(7S), S1–S34. https://doi.org/10.1016/j.cmi.2026.05.012

[24] Stoesser, N., Batty, E. M., Eyre, D. W., Morgan, M., Wyllie, D. H., Del Ojo Elias, C., et al. (2013). Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *Journal of Antimicrobial Chemotherapy*, 68(10), 2234–2244. https://doi.org/10.1093/jac/dkt180

[25] Timbrook, T. T., Morton, J. B., McConeghy, K. W., Caffrey, A. R., Mylonakis, E., & LaPlante, K. L. (2017). The effect of molecular rapid diagnostic testing on clinical outcomes in bloodstream infections: a systematic review and meta-analysis. *Clinical Infectious Diseases*, 64(1), 15–23. https://doi.org/10.1093/cid/ciw649

[26] Turnidge, J., Kahlmeter, G., Giske, C. G., Cantón, R., Matuschek, E., Wootton, M., et al. (2026). How to: the application and analysis of categorical agreement in antimicrobial susceptibility testing using European Committee on Antimicrobial Susceptibility Testing breakpoints. *Clinical Microbiology and Infection*, 32(6), 916–920. https://doi.org/10.1016/j.cmi.2025.12.018

[27] van Belkum, A., Bachmann, T. T., Lüdke, G., Lisby, J. G., Kahlmeter, G., Mohess, A., et al. (2019). Developmental roadmap for antimicrobial susceptibility testing systems. *Nature Reviews Microbiology*, 17(1), 51–62. https://doi.org/10.1038/s41579-018-0098-9

[28] van Belkum, A., Burnham, C. D., Rossen, J. W. A., Mallard, F., Rochas, O., & Dunne, W. M. (2020). Innovative and rapid antimicrobial susceptibility testing systems. *Nature Reviews Microbiology*, 18(5), 299–311. https://doi.org/10.1038/s41579-020-0327-x

[29] Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., et al. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, 155(8), 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

[30] Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158), 209–212. https://doi.org/10.1080/01621459.1927.10502953

[31] Zankari, E., Allesøe, R., Joensen, K. G., Cavaco, L. M., Lund, O., & Aarestrup, F. M. (2017). PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *Journal of Antimicrobial Chemotherapy*, 72(10), 2764–2768. https://doi.org/10.1093/jac/dkx217

Appendix A. Reproduction script

The script below downloads nothing itself; run it on the Europe PMC full-text XML (link in the header). It extracts source Tables 3 and 4, asserts the expected totals and prints every count, rate, interval and scenario value reported in Section 4.

```
"""Appendix A reproduction script: Stoesser et al. (2013) J Antimicrob Chemother 68:2234-44
(PMC3772739).
Usage: python -I p2_reproduce.py stoesser.xml
Download: https://www.ebi.ac.uk/europepmc/webservices/rest/PMC3772739/fullTextXML"""
import sys, re
from math import sqrt
from lxml import etree
Z = 1.959963984540054
def wilson(k, n):
    if n == 0: return (float('nan'),) * 2
    p = k / n; d = 1 + Z * Z / n; c = (p + Z * Z / (2 * n)) / d
    h = Z * sqrt(p * (1 - p) / n + Z * Z / (4 * n * n)) / d
    return max(0.0, c - h), min(1.0, c + h)
def extract(path):
    t = etree.parse(path); out = {}
    for tid, org in (("DKT180TB3", "E. coli"), ("DKT180TB4", "K. pneumoniae")):
        tw = [x for x in t.iter("table-wrap") if x.get("id") == tid][0]
        for tr in tw.iter("tr"):
            cells = ["".join(c.itertext()).strip() for c in tr]
            if len(cells) >= 5 and cells[0] not in ("Antibiotic", "Total") and re.match(r"^\d",
cells[1]):
                gs_cs, gr_cs, gs_cr, gr_cr = (int(c.split()[0]) for c in cells[1:5])
                out[(org, cells[0])] = (gs_cs, gr_cs, gs_cr, gr_cr)
    return out
def summarise(rows, label):
    ss = sum(r[0] for r in rows); rs = sum(r[1] for r in rows); sr = sum(r[2] for r in rows); rr =
sum(r[3] for r in rows)
    n = ss + rs + sr + rr; cr = sr + rr; cs = ss + rs
    print(f"{label}: n={n} agree={100*(ss+rr)/n:.2f}% CR={cr} CS={cs} "
          f"GS/CR={sr} ({100*sr/n:.2f}% all; {100*sr/cr:.2f}% of CR) "
          f"GR/CS={rs} ({100*rs/n:.2f}% all; {100*rs/cs:.2f}% of CS) sens={rr/cr:.4f}
spec={ss/cs:.4f}")
    return rr / cr, ss / cs
if __name__ == "__main__":
    d = extract(sys.argv[1])
    assert len(d) == 14
    for org, n in (("E. coli", 74), ("K. pneumoniae", 69)):
        assert all(sum(v) == n for (o, _), v in d.items() if o == org)
    for (o, drug), v in d.items():
```

```
    agree = (v[0] + v[3]) / sum(v); cr = v[2] + v[3]
    lo, hi = wilson(v[2], cr)
    print(f"{o:14s} {drug:14s} {v} agreement={100*agree:.2f}%",
          f"GS/CR={v[2]}/{cr} Wilson=({100*lo:.1f}, {100*hi:.1f})")
allrows = list(d.values())
assert sum(sum(v) for v in allrows) == 1001
perf = {}
for org in ("E. coli", "K. pneumoniae"):
    perf[org] = summarise([v for (o, _), v in d.items() if o == org], org)
summarise(allrows, "Combined")
kept = [v for k, v in d.items() if k != ("K. pneumoniae", "Amoxicillin")]
summarise(kept, "Excluding K. pneumoniae-amoxicillin")
for org, (se, sp) in perf.items():
    for p in (0.05, 0.50):
        post = (1 - se) * p / ((1 - se) * p + sp * (1 - p))
        print(f"Scenario {org} prevalence {p:.0%}:",
              f"P(comparator R | genomic S) = {100*post:.2f}%")
```