Electronic Journal of Medical Research
https://respublishers.com/journals/ejmr
ISSN: 3106-1362/3106-1370

Research article

When Does a Self-Assessment Carry Information? Regional Variation in the Corroboration of Reported Antimicrobial Resistance Surveillance Capacity Across 182 Countries

Aatiqa Masoom [1,2,*] and Hamna Masoom [3]

[1] Peking University Institute of Advanced Agricultural Sciences, Weifang 262113, China
[2] Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding, College of Agriculture, Yangzhou University, Yangzhou 225009, China
[3] School of Chemistry and Chemical Engineering, Yangzhou University, Yangzhou 225002, China

*Correspondence: masoomali663@yahoo.com

Article information:
Received: 04-10-2026
Revised: 01-11-2026
Accepted: 01-11-2026
Published: 03-11-2026

Academic Editor:
Dr. Anwar Althubyani

Keywords:
Antimicrobial resistance
Self-assessment
Measurement validity
Ceiling effect
Health governance
TrACSS

Abstract: International antimicrobial resistance (AMR) monitoring relies on the unverified Tracking AMR Country Self-Assessment Survey (TrACSS). Whether its ratings provide equivalent information across settings remains unclear. This cross-sectional analysis examined correspondence between self-rated surveillance capacity and laboratory capabilities in the 2025 public dataset, by WHO region and income. For 182 countries answering surveillance item 3.5, a corroboration score (0–4) incorporated reference-laboratory susceptibility testing for 12 priority bacteria, standardized testing guidelines and external quality assurance beyond reference laboratories, and reporting a laboratory count. Associations were assessed using Spearman correlations. Mean corroboration increased from 1.50 at rating A to 3.43 at rating D (overall $\rho = 0.43$; 95% CI 0.29–0.55; $p < 0.001$). Regional correlations ranged from 0.67 in the Eastern Mediterranean to 0.20 in the Western Pacific; confidence intervals excluded zero in three regions but included zero in Europe and the Western Pacific. Mean corroboration did not differ significantly between regions (Kruskal–Wallis $p = 0.80$). Across six regions, correlation strength was inversely rank-ordered with corroboration among low self-raters ($\rho = -1.00$; exact $p = 0.003$), consistent with ceiling effects. Among 107 countries claiming standardized national surveillance, 68 (63.6%) reported action plans at best partially implemented. Laboratory-count non-reporting increased with income, from 7.4% to 27.8% (trend $p = 0.035$). Self-ratings showed moderate overall correspondence with laboratory capabilities reported by the same respondents, with regional variation. These findings support reporting corroboration alongside ratings and caution against assuming equivalent measurement performance across settings; they do not constitute independent validation of national capacity.

Article citation: Masoom A. & Masoom, H. When Does a Self-Assessment Carry Information? Regional Variation in the Corroboration of Reported Antimicrobial Resistance Surveillance Capacity Across 182 Countries. Electron J Med Res. 2026. 2(4): 111-118. https://doi.org/10.64813/ejmr.2026.156

Introduction

The Tracking AMR Country Self-Assessment Survey is the principal instrument through which the international community observes national progress against antimicrobial resistance. Administered annually by the Quadripartite organizations and completed by each country's national AMR focal point, it supplies the indicators of the monitoring and evaluation framework of the Global Action Plan on AMR, and the 2024 United Nations General Assembly political declaration sets a target that 95% of countries respond each year [1]. The

Copyright © 2026 Masoom A. & Masoom, H. This work is published and licensed by RES Publishers.
The full terms of this license are available at https://www.respublishers.com/terms_conditions.php.

ninth round, conducted in 2025, received 184 country responses [2].

Two substantial analyses published in 2026 built upon these data. Chen and colleagues constructed a One Health governance index across 193 countries and found that implementation lagged behind policy design [3]. Drake and colleagues examined surveillance governance across two survey rounds and assessed the effect of a major donor programme. Both noted the instrument's dependence on self-report, and Drake and colleagues recorded that "no effort was made to triangulate or otherwise validate TrACSS data as part of the analysis," identifying external triangulation as a task for others [4-6].

Validation is usually framed as a single question: are the responses accurate? That framing may be the wrong one. A self-assessment can be informative in one setting and uninformative in another without any country misreporting, because the information a rating conveys depends on how much variation there is to convey [7, 8]. Where national capacity differs widely between countries, a rating that tracks capacity is highly informative. Where capacity is uniformly adequate, the same rating distinguishes little, and the residual variation it captures may be stylistic rather than substantive [9, 10].

This distinction matters for how TrACSS is used. Regional comparisons, donor targeting and global progress narratives all treat the ratings as commensurable across settings. If their information content varies systematically, then the same numerical difference means different things in different places, and some comparisons that appear straightforward are not [4, 11]. We therefore asked three questions of the complete 2025 round. First, does a country's self-rated national AMR surveillance capacity correspond to the laboratory capacity that country separately reports? Second, does the strength of that correspondence vary by region and by income group? Third, if it varies, what accounts for the variation? The analysis deliberately stays inside the dataset: it tests internal corroboration rather than external accuracy, which is a weaker test but one that requires no additional data and that no prior study has performed.

Methods

Design and data

Cross-sectional analysis of the TrACSS 2025 public data export, released on 1 September 2025 and covering the ninth survey round. The export contains 184 country records across 248 variables. The questionnaire was issued on 15 April 2025 with a deadline of 15 June 2025 and requests a single consolidated national response coordinated by the national AMR focal point. Ethical approval was not required: the data are publicly released, aggregated to country level, and contain no personal information.

Self-rated surveillance capacity

The exposure was item 3.5, national surveillance system for antimicrobial resistance (AMR) in humans, rated A to E. A denotes no capacity to generate or report AMR data; B, local collection without necessarily using a standardised approach; C, national collection without necessarily using a standardised approach; D, a standardised national system with a national coordinating centre collecting data on common bacterial infections in hospitalised and community patients; and E, as D with additional linkage to antimicrobial consumption or use data. Ratings of D or E assert a standardised, nationally coordinated system. Of the 184 responding countries, 182 answered this item and constitute the analysis population.

Corroboration score

Four further items, answered by the same respondent, were specified in advance as the minimum laboratory substrate implied by a standardised national surveillance system. One point was assigned for each of the following: a national bacteriology or AMR reference laboratory performing susceptibility testing for all 12 listed priority bacteria (item 3.7); standardised susceptibility testing guidelines in use by at least some clinical laboratories rather than the reference laboratory alone (item 3.11); an external quality assurance programme extending beyond the national reference laboratory (item 3.12); and a reported numerical count of public-sector laboratories performing culture and susceptibility testing (item 3.6).

The fourth item measures reporting rather than capacity, and is included deliberately. A country operating a coordinated national surveillance system should be able to state how many laboratories contribute to it; a blank response is therefore treated as a failure of corroboration rather than as missing data to be imputed. Scores range from 0 to 4.

Analysis

The association between self-rating, treated as an ordinal variable from 1 to 5, and corroboration score was assessed by Spearman rank correlation, computed overall and within each WHO region. Confidence intervals were derived by Fisher z transformation with

the standard error 1.06/√(n−3) appropriate to Spearman coefficients. Differences in corroboration score across regions and across World Bank income groups were tested by Kruskal–Wallis analysis of variance. The relationship between regional correlation strength and regional baseline capacity was assessed by Spearman correlation across the six regions, with significance determined by exact permutation over all 720 orderings rather than an asymptotic approximation, given the small number of units.

Secondary comparisons used item 2.3.a, country progress with development of a national action plan on AMR, on the same five-point scale, together with item 2.29, whether an integrated AMR surveillance system has been established. Analyses used Python 3.13 with SciPy 1.18. The extraction script and coded dataset are supplied so that every value reported here can be reproduced from the published export.

Results

Corroboration rises with self-rating

Across all 182 countries, mean corroboration rose monotonically with self-rated capacity, from 1.50 among the eight countries rating A to 3.02 at C and 3.43 at D. The overall association was moderate and clearly non-zero (Spearman ρ = 0.43, 95% CI 0.29 to 0.55; p < 0.001). Respondents are therefore not answering arbitrarily: on average, countries claiming more also report more of the laboratory infrastructure that such claims require. Results by rating level are given in Table 1.

One departure from monotonicity is notable. Countries rating E, the highest available rating, recorded a mean corroboration of 3.41, marginally below the 3.43 of those rating D. The top of the scale does not distinguish itself from the rung below on any of the four laboratory criteria, which is the first indication of a ceiling.

The association varies by region

Disaggregated by WHO region, the association ranged more than threefold. It was strongest in the Eastern Mediterranean Region (ρ = 0.67, 95% CI 0.28 to 0.87; p = 0.002), followed by the Americas (ρ = 0.64, 0.36 to 0.81; p < 0.001) and Africa (ρ = 0.50, 0.23 to 0.70; p < 0.001). In Europe (ρ = 0.22, −0.09 to 0.49; p = 0.136) and the Western Pacific Region (ρ = 0.20, −0.22 to 0.56; p = 0.321) the confidence interval included zero. The South-East Asia Region produced a point estimate of 0.57 but with only nine responding countries the interval was wide and included zero. Regional results are shown in Figure 1 and Table 2.

Europe and the Western Pacific contributed 48 and 26 countries respectively, so their null results are not simply a matter of power. In those two regions, knowing a country's self-rating conveys essentially nothing about whether its own reported laboratory capacity supports that rating.

The variation is not a difference in capacity

An obvious explanation would be that corroboration is simply higher in some regions than others. It is not. Mean corroboration score did not differ significantly across the six regions (Kruskal–Wallis H = 2.32, p = 0.80), and the regional means clustered narrowly. Nor was income group decisive: mean corroboration was 3.37 in high-income, 2.90 in upper-middle-income, 3.22 in lower-middle-income and 2.93 in low-income countries, a pattern that is neither monotonic nor statistically significant (H = 6.44, p = 0.092).

What differs is not the level of corroboration but the extent to which the self-rating tracks it.

Table 1. Corroboration of self-rated national AMR surveillance capacity, 182 responding countries, 2025.

Self-rating on item 3.5	Countries (n)	Mean criteria met	Met all four, n (%)
A — no capacity	8	1.50	0 (0%)
B — local collection	24	2.50	3 (12%)
C — national, not standardised	43	3.02	11 (26%)
D — standardised national system	90	3.43	48 (53%)
E — as D, plus consumption linkage	17	3.41	8 (47%)
All countries	182	3.13	70 (38%)

Criteria met counts the four corroborating items defined in the Methods, each reported by the same respondent. Corroboration rises monotonically from rating A to rating D but does not increase further at rating E. Source: TrACSS 2025 data export.

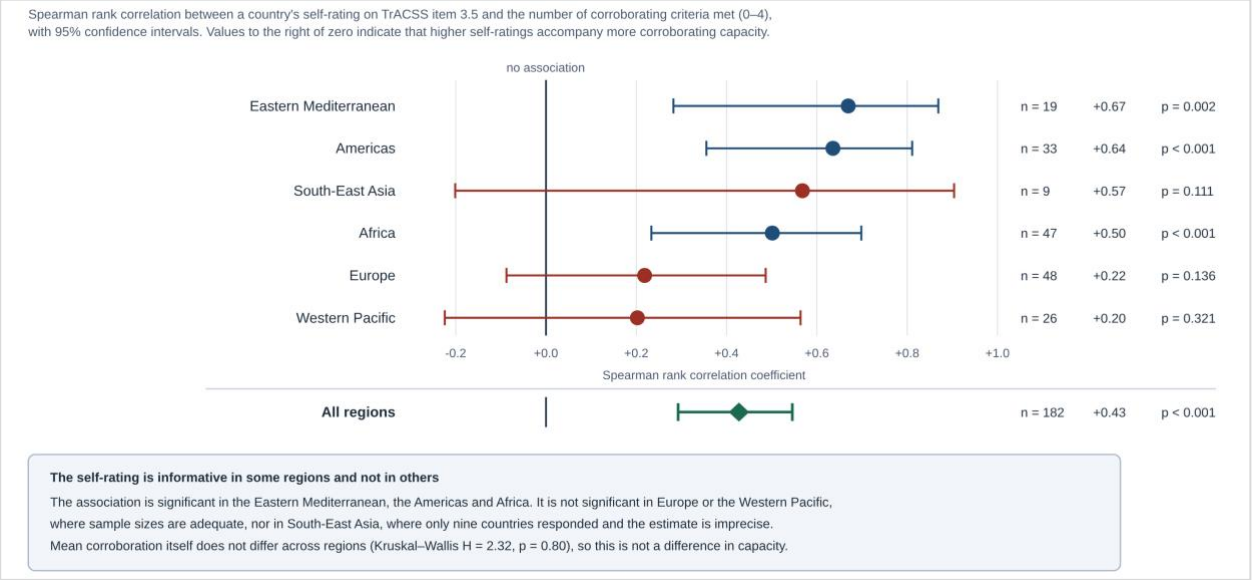


Figure 1. Association between self-rated surveillance capacity and its corroboration, by WHO region. Spearman rank correlation between a country's self-rating on TrACSS item 3.5 and the number of corroborating criteria met, with 95% confidence intervals. The association is significant in the Eastern Mediterranean, the Americas and Africa; the confidence interval includes zero in Europe and the Western Pacific, where sample sizes are adequate, and in South-East Asia, where only nine countries responded. Mean corroboration does not differ across regions (Kruskal–Wallis H = 2.32, p = 0.80). Source: TrACSS 2025 data export, WHO/FAO/UNEP/WOAH.

Table 2. Association between self-rated surveillance capacity and corroboration, by WHO region.

WHO region	Responding countries	Rating D or E, n (%)	Spearman ρ (95% CI)	p	Mean criteria, A–C	Mean criteria, D–E	Gap
Eastern Mediterranean	19	14 (73.7%)	+0.67 (+0.28 to +0.87)	0.002	1.60	3.50	+1.90
Americas	33	17 (51.5%)	+0.64 (+0.36 to +0.81)	<0.001	2.44	3.53	+1.09
South-East Asia	9	6 (66.7%)	+0.57 (-0.20 to +0.90)	0.111	2.67	3.50	+0.83
Africa	47	24 (51.1%)	+0.50 (+0.23 to +0.70)	<0.001	2.74	3.38	+0.64
Europe	48	32 (66.7%)	+0.22 (-0.09 to +0.49)	0.136	2.94	3.38	+0.44
Western Pacific	26	14 (53.8%)	+0.20 (-0.22 to +0.56)	0.321	3.08	3.43	+0.35
All regions	182	107 (58.8%)	+0.43 (+0.29 to +0.55)	<0.001	2.69	3.43	+0.74

Spearman rank correlation between self-rating on item 3.5 (treated as ordinal, 1 to 5) and corroboration score (0 to 4). Confidence intervals by Fisher z transformation. Gap is the difference in mean criteria met between countries rating D or E and those rating A to C. Mean corroboration does not differ significantly across regions (Kruskal–Wallis H = 2.32, p = 0.80). Source: TrACSS 2025 data export.

A ceiling effect explains the pattern

Separating each region's countries into low raters (A to C) and high raters (D or E) locates the difference. Among high raters, mean corroboration was almost identical everywhere, ranging only from 3.38 to 3.53. Among low raters it varied widely, from 1.60 in the Eastern Mediterranean to 3.08 in the Western Pacific. In other words, regions differ in how much capacity their low-rating countries have, not in how much their high-rating countries have.

The consequence is arithmetical. Where low raters already report substantial laboratory capacity, as in the Western Pacific, there is little room between low and high raters for the rating to discriminate, and the correlation approaches zero. Where low raters genuinely lack that capacity, as in the Eastern Mediterranean, the rating separates them sharply.

This relationship is exact. Ordering the six regions by correlation strength produces precisely the reverse of their ordering by baseline corroboration among low raters: Spearman ρ = −1.00 across the six regions, exact p = 0.003 by permutation over all 720 orderings. With six units this is a descriptive rather than an inferential result, but a perfect inversion is a strong indication that the mechanism is a ceiling effect rather than a difference in reporting behaviour. Figure 2A presents the relationship.

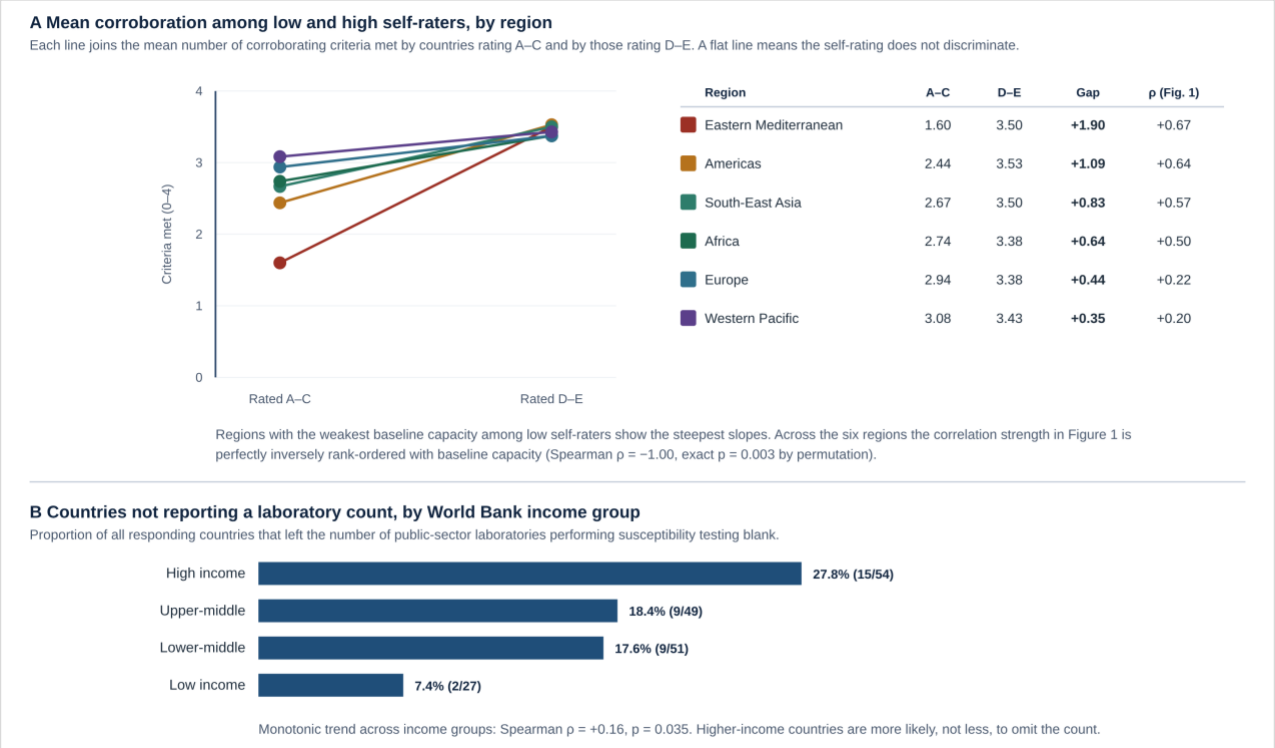


Figure 2. Why the association varies. (A) Mean corroboration among countries rating A to C and those rating D or E, by region. High raters cluster narrowly between 3.38 and 3.53 in every region, while low raters range from 1.60 to 3.08, so the room available for the rating to discriminate differs by setting. Across the six regions, the correlation strength shown in Figure 1 is perfectly inversely rank-ordered with baseline corroboration among low raters (Spearman ρ = −1.00; exact p = 0.003 by permutation over all 720 orderings). (B) Proportion of responding countries leaving the laboratory count blank, by World Bank income group, showing a monotonic trend in the direction opposite to a resource explanation. Source: TrACSS 2025 data export, WHO/FAO/UNEP/WOAH.

Two patterns that are regionally general

Two further findings held across regions. First, of the 107 countries rating their surveillance system D or E, 68 (63.6%) simultaneously reported that their national action plan on AMR was at best partially implemented with limited resources. Regional proportions ranged from 47% in Europe to 83% in South-East Asia. Forty-four of the 107 (41%) also reported that no integrated AMR surveillance system had been established or begun implementation. A standardised national surveillance system is among the more resource-intensive deliverables of a national action plan, and these positions are not easily reconciled. Results are given in Table 3.

Second, failure to report a laboratory count rose monotonically with national income: 7.4% of low-income countries left the item blank, against 17.6% of lower-middle-income, 18.4% of upper-middle-income and 27.8% of high-income countries (Spearman ρ = +0.16 across income ranks, p = 0.035). Among countries claiming a standardised national system, 29 of 107 (27.1%) did not report a count. This is the opposite of a resource explanation, under which the least-resourced countries would be least able to enumerate their networks. Figure 2B presents the gradient.

Discussion

The question of whether a self-assessment is valid admits a more useful reformulation: for which comparisons is it valid? These data suggest that TrACSS surveillance ratings are informative in the Eastern Mediterranean, the Americas and Africa, and essentially uninformative in Europe and the Western Pacific, not because countries in the latter regions report less honestly, but because there is less for the rating to distinguish [1, 2, 10]. The practical implication is specific. Comparative claims that rank high-capacity regions or countries against one another on the basis of these ratings are not supportable, because within those groups the rating does not correspond to reported capacity. Comparative claims within lower-capacity regions, and assessments of change over time in those regions, rest on firmer ground. This is a more optimistic conclusion than a blanket caution about self-report,

and a more demanding one, because it requires users of the data to establish which case they are in rather than treating the instrument as uniformly reliable.

Table 3. Discordant responses among the 107 countries rating their national AMR surveillance system as a standardised national system or better.

WHO region	Countries rating D or E	NAP at C or below, n (%)	No integrated system, n (%)	Laboratory count not reported, n (%)
Eastern Mediterranean	14	11 (79%)	5 (36%)	5 (36%)
Americas	17	12 (71%)	6 (35%)	1 (6%)
South-East Asia	6	5 (83%)	2 (33%)	3 (50%)
Africa	24	17 (71%)	11 (46%)	5 (21%)
Europe	32	15 (47%)	12 (38%)	10 (31%)
Western Pacific	14	8 (57%)	8 (57%)	5 (36%)
All regions	107	68 (64%)	44 (41%)	29 (27%)

National action plan rating is item 2.3.a; C or below denotes a plan that is, at best, partially implemented with limited resources. Integrated surveillance system is item 2.29. Laboratory count refers to the number of public-sector laboratories performing culture and susceptibility testing, item 3.6. Source: TrACSS 2025 data export.

Chen and colleagues reported that implementation lags policy design globally [7]. The present analysis identifies the same divergence within a single questionnaire round and at the level of individual responses: countries' claims about surveillance systems outrun their claims about the plans that fund those systems, in nearly two-thirds of cases. That this holds across all six regions suggests it is a property of how capacity is reported rather than a feature of any particular setting.

Drake and colleagues called for triangulation against external sources and did not undertake it [4]. This study does not supply that triangulation; it establishes something prior and arguably more consequential, which is that the instrument's information content is not constant across the settings being compared. External validation against observed surveillance output remains necessary, and would be most informative in precisely the high-capacity regions where internal corroboration is least able to discriminate.

The monotonic income gradient in unreported laboratory counts invites explanation. One possibility is structural: in countries with large private laboratory sectors and decentralised health systems, the number of public-sector laboratories performing susceptibility testing may be genuinely difficult to establish, and may be a less meaningful quantity than in systems where a defined network reports to a national centre [8, 9]. Another is procedural: focal points in well-resourced systems may complete the ordinal items and omit the free-text counts, which require consultation across institutions. The data cannot distinguish these, but either reading implies that the item does not measure the same thing in all settings, and that cross-national comparison of laboratory counts should be avoided.

Three changes would improve interpretability without altering the questionnaire's purpose or burden. First, published country responses could carry a corroboration indicator computed from the instrument's own internal logic, so that a rating of D accompanied by no external quality assurance programme and no laboratory count is visible as such. Second, regional summaries should report response rates, since a regional figure computed on a subset of member states is not a regional figure. Third, the laboratory count item requires an explicit definition; the range of reported values and the income gradient in non-response both suggest that respondents are not answering the same question. More broadly, the finding argues against the construction of composite governance indices that pool TrACSS responses across regions without accounting for differential information content [1]. An index that weights a European rating and an Eastern Mediterranean rating equally is combining a measure that discriminates with one that does not.

The corroboration criteria were specified by the authors rather than derived from a validated framework. They were fixed before analysis and the coded dataset is supplied so that alternative specifications can be applied, but a different reasonable choice would yield somewhat different coefficients. The ceiling-effect interpretation does not depend on the specific criteria, only on the observation that high raters cluster near the maximum. Internal corroboration cannot detect uniform overstatement. A country that

overstates every item consistently scores four of four. The analysis therefore provides a lower bound on incoherence, not an estimate of inaccuracy, and the regions in which the rating appears informative may still contain systematically optimistic responses [2, 10].

Discordance between a surveillance claim and a national action plan rating is not proof of error. Surveillance systems are sometimes financed through donor or vertical programs outside the national plan, so both answers may be correct. The 68 discordant responses identified here require explanation rather than constituting demonstrated misstatements. The perfect rank inversion between correlation strength and baseline capacity rests on six units. It is a striking descriptive pattern, and the exact permutation p-value is reported in preference to an asymptotic one, but six regions cannot establish a mechanism on their own, and replication across further survey rounds would strengthen the inference considerably.

Finally, these are data from one round. The questionnaire was revised for round nine in the multisector, human health and environment sections, so whether the patterns described are stable over time cannot be determined here.

Conclusion

Self-rated national AMR surveillance capacity corresponds to reported laboratory capacity overall, but the strength of that correspondence varies more than threefold between WHO regions, and in Europe and the Western Pacific it is indistinguishable from zero. Mean corroboration does not differ between regions, so this is a difference in what the rating conveys rather than in what countries have. The explanation appears to be a ceiling effect: the rating discriminates where capacity is heterogeneous and fails to discriminate where it is uniformly adequate. The instrument is therefore neither reliable nor unreliable in general; its usefulness depends on the comparison being made. Users should establish which case applies before drawing conclusions, report corroboration alongside the ratings, and resist composite indices that assume equivalent information content across settings. Two further findings warrant attention in their own right: nearly two-thirds of countries claiming a standardised national surveillance system report national action plans they describe as not fully implemented, and the countries least likely to enumerate their laboratory networks are the wealthiest.

Author Contributions: Conceptualization, A.M.; methodology, A.M.; software, A.M.; validation, H.M.; formal analysis, A.M.; investigation, H.M.; resources, H.M.; data curation, A.M.; writing—original draft preparation, A.M.; writing—review and editing, H.M.; visualization, A.M.; supervision, A.M.; project administration, A.M. Both authors have read and agreed to the final version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. The study analyzed publicly accessible web content and involved no human participants or personal data.

Informed Consent Statement: Not applicable.

Data Availability Statement: The primary data are the TrACSS 2025 public data export, available from the Quadripartite AMR country progress portal. Additional data will be made available upon reasonable request to the corresponding author.

Acknowledgments: The authors acknowledge the Peking University Institute of Advanced Agricultural Sciences for facilitating this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Quadripartite Joint Secretariat on Antimicrobial, R., Tracking AMR Country Self-assessment Survey (TrACSS) 9.0, 2025: cover letter, introduction and questionnaire. **2025**.
2. Agboola, O.E., Agboola, S.S., et al. Global surveillance quality gaps in antimicrobial resistance monitoring: a composite-index and inequality analysis of 85 countries. JAC-Antimicrobial Resistance. **2026**. 8(4): dlag146.
3. World Health, O., Global antibiotic resistance surveillance report 2025: WHO Global Antimicrobial Resistance and Use Surveillance System (GLASS). **2025**, World Health Organization: Geneva.
4. Drake, A., Sassoon, I., et al. Country governance of antimicrobial resistance (AMR) surveillance: observations on global progress and aid programme effectiveness using data from the Tracking AMR Country Self-Assessment Survey (TrACSS). Globalization and Health. **2026**. 22(1): 8 10.1186/s12992-025-01179-4.
5. Alhusseini, L.B., Hasani, B., et al. Temporal and geographic trends in extended-spectrum cephalosporins resistance among Neisseria gonorrhoeae isolates worldwide: a systematic review and meta-analysis. BMC Infectious Diseases. **2025**. 25(1): 1175.
6. Frost, I., Kapoor, G., et al. Status, challenges and gaps in antimicrobial resistance surveillance around the

world. Journal of global antimicrobial resistance. **2021**. 25: 222–226.
7. Chen, W., Zeng, Y., et al. Evaluation of antimicrobial resistance governance across 193 countries to inform the 2026 Global Action Plan update. Nature Medicine. **2026**. 32(4): 1362–1373 10.1038/s41591-026-04257-1.
8. Mestrovic, T., Naghavi, M., et al. The burden of bacterial antimicrobial resistance in the WHO Eastern Mediterranean Region 1990–2021: a cross-country systematic analysis with forecasts to 2050. The Lancet Public Health. **2025**. 10(11): e955–e970.
9. World Health, O. and Our World in, D. Countries enrolled in GLASS surveillance of antimicrobial consumption and resistance. **2024**.
10. Tosas Auguet, O., Dean, A., et al. Global antibiotic resistance surveillance report 2025. **2025**.
11. Sartorius, B., Gray, A.P., et al. The burden of bacterial antimicrobial resistance in the WHO African region in 2019: a cross-country systematic analysis. The Lancet Global Health. **2024**. 12(2): e201–e216.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of RES Publishers and/or the editor(s). RES Publishers and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.