



Original article

External validation of WGS-based antimicrobial susceptibility prediction tools, KOVER-AMR and ResFinder 4.1, for *Escherichia coli* clinical isolates

Tess Verschuuren^{1,*}, Thijs Bosch², Valentina Mascaro³, Rob Willems⁴, Jan Kluytmans^{1,5}

¹ Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, the Netherlands

² The Netherlands Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

³ ASST Rhodense, G. Salvini Hospital, Garbagnate Milanese, Italy

⁴ Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, the Netherlands

⁵ Amphia Hospital Breda, Breda, the Netherlands

ARTICLE INFO

Article history:

Received 30 October 2021

Received in revised form

8 May 2022

Accepted 14 May 2022

Available online 2 June 2022

Editor: E. Yusuf

Keywords:

E. coli

KOVER-AMR

ResFinder

WGS-AST

External validation

ABSTRACT

Objective: To externally validate whole genome sequence-antimicrobial susceptibility testing phenotype prediction tools KOVER-AMR and ResFinder 4.1 for *Escherichia coli* clinical isolates from Dutch routine care. **Methods:** A random sample of 234 *E. coli* and 283 third generation cephalosporin-resistant *E. coli* isolates from urine and blood were collected (2014–2017). Culture-antimicrobial susceptibility testing was performed using VITEK 2 and BD Phoenix. Sequences were used as input for KOVER-AMR-SCM, KOVER-AMR-CART, and ResFinder 4.1. The concordance, major error rate (MER), and very major error rate (VMER) were calculated, with subsequent comparison to U.S. Food and Drug Administration (FDA) criteria (MER $\leq 3\%$ and VMER with a 95% confidence interval $\leq 1.5\text{--}7.5\%$).

Results: ResFinder 4.1 performed better than KOVER-AMR-models; however, neither tool achieved overall (V)MERs below FDA criteria. KOVER-AMR-SCM, KOVER-AMR-CART, and ResFinder 4.1, MER (cumulative all antimicrobials) were: 5.1% (4.4–5.9), 4.3% (3.6–5.0), and 5.1% (4.5–5.8), respectively. MERs $\leq 3\%$ were achieved for 6 (SCM) and 5 (CART) of the 11 tested antimicrobials for KOVER-AMR-models and for 9/13 antimicrobials tested with ResFinder 4.1. KOVER-AMR-SCM, KOVER-AMR-CART, and ResFinder 4.1 cumulative VMERs were: 26% (24–28), 29% (27–31), and 11% (9.2–12). VMERs with a 95% CI $\leq 1.5\text{--}7.5\%$ were only achieved for 4/13 tested antimicrobials with ResFinder 4.1.

Discussion: In this study, whole genome sequence-antimicrobial susceptibility testing phenotype prediction tools KOVER-AMR and ResFinder 4.1 did not meet the FDA criteria needed for clinical diagnostic use in 517 *E. coli* clinical isolates from Dutch routine care. The tested tools should be further improved before they can be used for clinical decision making. **Tess Verschuuren, Clin Microbiol Infect 2022;28:1465**

© 2022 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Escherichia coli is the most important cause of urinary tract-, bloodstream-, and antimicrobial-resistant infections in Europe [1,2]. Culture-based antimicrobial susceptibility testing (AST) (culture-AST) is performed to provide adequate treatment for *E. coli* infections [3,4]. Culturing followed by AST is affordable but takes 2 to 3 days, as the process depends on bacterial growth. Awaiting

culture-AST results, a patient is treated with empiric broad-spectrum therapy. As a consequence, patients can be temporarily overtreated (resulting in unnecessary antimicrobial use) or undertreated (resulting in potential adverse patient outcomes).

Whole genome sequencing (WGS) could provide the possibility to replace the practice of empiric therapy with point-of-care AST-directed therapy and solve issues of between-lab comparability, reanalysis, and storage [5–8]. For this, the following developments are needed in clinical microbiology: (i) high throughput sequencing and analyses workflows [5]; (ii) development of methodologies that allow sequencing directly on a clinical sample, especially on

* Corresponding author. Tess Verschuuren.

E-mail address: t.d.verschuuren@gmail.com (T. Verschuuren).

material that is sterile under normal conditions (e.g. blood) [9–11]; and (iii) WGS-based predictive AST (WGS-AST) tools that provide direct phenotype predictions (e.g. amoxicillin resistance yes/no), without interpretation of genomic content (e.g. presence of *bla*_{TEM-1}), which will save time and prevent interpretation errors [12]. WGS-AST technologies should adhere to quality criteria set by the U.S. Food and Drug Administration (FDA). These are expressed as a major error rate (MER) of $\leq 3\%$ (i.e. susceptible isolates falsely predicted as resistant, resulting in overtreatment), and a very major error rate (VMER) including an upper 95% CI limit of ≤ 7.5 for the true VMER and a lower 95% CI limit of ≤ 1.5 , further referred to as 95% CI of ≤ 1.5 – ≤ 7.5 (i.e. resistant isolates falsely predicted as susceptible, resulting in undertreatment) [13].

Studies have shown promising results of WGS-AST for several pathogens [14,15]. Unfortunately, (V)MERs are infrequently reported, and external validation studies are scarce [12,14,16]. KOVER-AMR [17] and ResFinder 4.1 [7] are two tools that provide direct phenotype predictions. Both tools observed a high concordance with culture-AST in the original studies of 95% (KOVER-AMR) and 98% (ResFinder 4.1) [7,17]. Here, we present an external validation of KOVER-AMR and ResFinder 4.1 using *E. coli* isolates recovered from infections in Dutch routine clinical care.

Methods

Study design

Sample collection of included isolates was described previously [18,19]. In short, a random sample of 234 *E. coli* isolates from blood (~1/3 of available unique isolates), and 283 third-generation cephalosporin-resistant (3rdGCR)-*E. coli* from urine and blood

were included from the years 2014 to 2017 from two hospitals and one primary care laboratory: the University Medical Center Utrecht (UMCU), a tertiary centre; the Amphia hospital, a large teaching centre; and Saltro, a laboratory providing services to primary care in the Utrecht region.

Ethics

The study was judged outside the scope of the Medical Research Involving Human Subjects Act by the ethics review board of the UMCU (IRB number 18/056). Informed consent was not obtained based on the 'Code of conduct for health research' [20].

Culture-AST

Phenotypes were determined using VITEK 2 (Amphia hospital and Saltro) and BD Phoenix (UMCU) and extracted for 14 antimicrobials (Fig. 1) [3]. Minimum inhibitory concentrations (MICs) were converted to susceptible or resistant based on the EUCAST clinical MIC breakpoints for *Enterobacterales* (v9.0), intermediate results were considered resistant [21]. VITEK 2 ADVANCED EXPERT SYSTEM™ rules were ignored. In consultation with a medical microbiologist, the ceftriaxone phenotypes from the UMCU were relabelled as cefotaxime, based on their similar working profile. In 11 of the 517 cultures (all blood) > 1 phenotypically distinct antimicrobial resistance *E. coli* isolates were retrieved. For these cultures, the isolate with the phenotype that corresponded with the WGS-AST predictions was selected for further analysis.

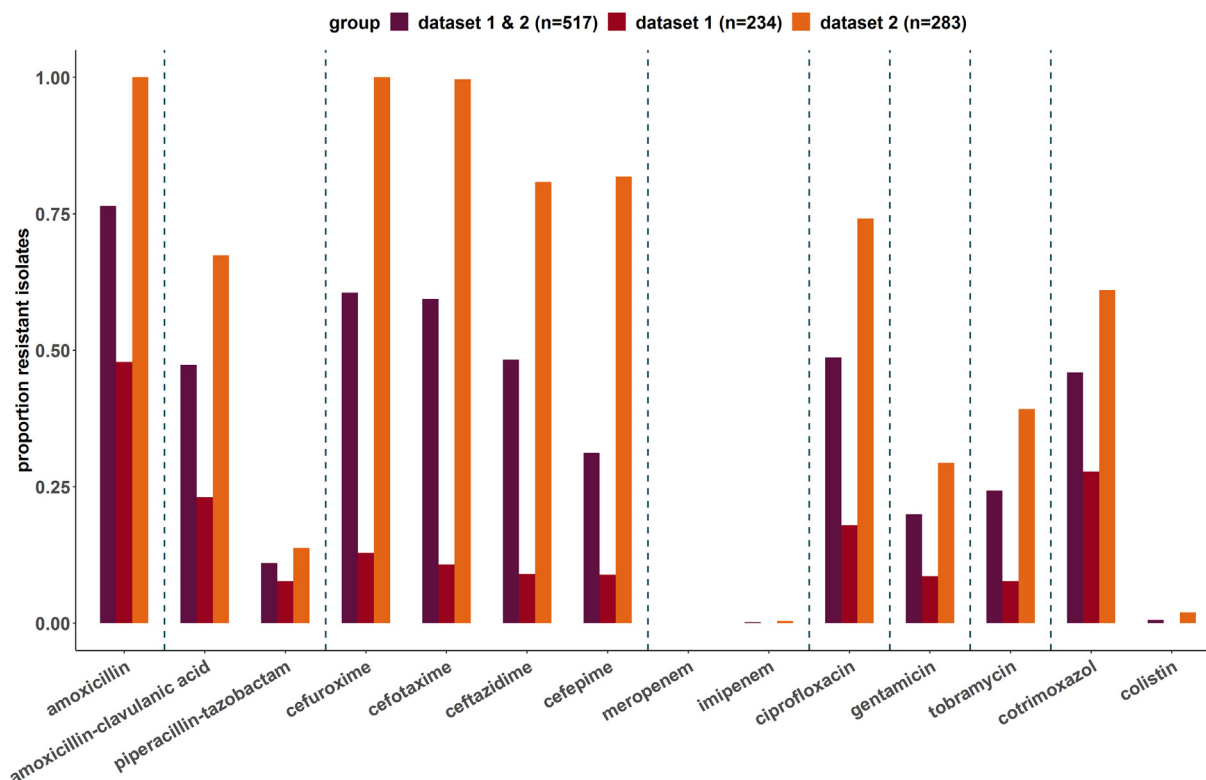


Fig. 1. Observed phenotypic resistance proportions of included antimicrobials. Dataset 1: a random sample of available *E. coli* isolates retrieved from blood from two hospitals. Dataset 2: Third-generation cephalosporin-resistant *E. coli* retrieved from urine and blood from two hospitals and one primary care laboratory. Dashed lines indicate antimicrobial classes, from left to right: penicillins (small-spectrum), penicillins (broad-spectrum), cephalosporins, carbapenems, aminoglycosides, miscellaneous agents. Missing phenotypes: 1 for amoxicillin-clavulanic acid, 1 for ceftazidime, 193 for cefepime, 1 for meropenem, 30 for imipenem, 23 for tobramycin, 1 for cotrimoxazol, and 183 for colistin.

WGS

WGS was performed using the Illumina HiSeq 2500 platform and the quality of the *de novo* assemblies was assessed (Appendix S1, and Supplementary material 2) [18,19]. Multilocus sequence types were determined, and core-genome neighbour joining trees were constructed to demonstrate the diversity of the included isolates (Appendix S2).

WGS-AST

KOVER, a freely available, supervised machine learning algorithm, was developed and used by Drouin et al. to produce two rule-based models for WGS-AST phenotype prediction: KOVER-AMR-Set Covering Machines (SCM) and KOVER-AMR-Classification and Regression Trees (CART), using a public genotype-phenotype database, based on the presence or absence of certain *k*-mers (a string of DNA with the length of *k*) [17,22]. Models are available for 12 bacterial species, for 56 antimicrobial treatment options [17]. Currently, no tool or package exists of KOVER-AMR. Available *E. coli* models were accessed at https://github.com/aldro61/kb_kover_amr/tree/master/data/models and imported in R Studio into a custom R-script (Appendix S3).

Bortolaia et al. published an update of ResFinder, a freely available and centrally curated reference database of acquired resistance genes and point mutations [7]. Including a genotype-to-phenotype translation for each resistance predictor in the database.

ResFinder 4.1 was accessed at <https://cge.cbs.dtu.dk/services/ResFinder/>. Assemblies were uploaded with the following default settings: (i) search for chromosomal point mutations and (ii) search for acquired antimicrobial resistance genes in all antimicrobial classes, both with a %ID of $\geq 90\%$ and a minimal length of $\geq 60\%$. The difference in number of misclassifications using an increased minimal length were visualised (Appendix S4).

Statistical analyses

The resulting output of both tools was a prediction of resistance (yes/no) for a particular antimicrobial per isolate. With this, the concordance, MER, and VMER were calculated (formulas in Appendix S5). 95% CIs were calculated with bootLR package (v1.0.2) or a test of given proportions. Calculations were performed with R Studio. MERs and VMERs were compared to FDA criteria (MER $\leq 3\%$ and VMER with a 95% CI of $\leq 1.5\text{--}7.5\%$).

Results

Phenotypic resistance rates of the included isolates ranged from 0% for meropenem to 76% for amoxicillin (Fig. 1). ResFinder 4.1 performed better than KOVER-AMR-models and provided predictions for more of the 14 assessed antimicrobials (13 vs 11) (Table 1). However, neither of the tools achieved overall (V)MERs below FDA criteria, with only ResFinder 4.1 predictions for

Table 1

The diagnostic performance of KOVER-AMR-SCM, KOVER-AMR-CART, and ResFinder 4.1 for the phenotypic antimicrobial susceptibility pattern of *E. coli* UTI and bacteraemia (*n* = 517) from two Dutch hospitals and one primary care laboratory

A ^a	Concordance % (95% CI)			ME rate % (95% CI)			VME rate % (95% CI)		
	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1
AMO (SSP)	75 (71–79)	75 (71–79)	98 (97–99)	3.3 (1.3–8.1)	3.3 (1.3–8.1)	1.6 (0.5–5.8)	32 (27–36)	32 (27–36)	1.8 (0.9–3.6)
AMC	72 (68–76)	75 (71–79)	72 (68–76)	3.7 (2.0–6.7)	5.9 (3.7–9.3)	2.6 (1.3–5.2)	55 (49–61)	47 (41–53)	57 (50–63)
PITA	84 (81–87)	84 (81–87)	81 (78–84)	7.0 (5.0–9.7)	7.0 (5.0–9.7)	16 (13–20)	88 (77–94)	88 (77–94)	39 (27–52)
BRSP	78 (75–81)	79 (77–82)	76 (74–79)	5.7 (4.4–7.7)	6.6 (5.0–8.6)	11 (9.1–14)	61 (56–67)	55 (49–60)	53 (48–59)
CER	96 (94–98)	96 (94–98)	NA ^b	0 (0–1.8)	0 (0–1.8)	NA ^b	6.4 (4.2–9.7)	6.4 (4.2–9.7)	NA ^f
CFO	97 (95–98)	97 (95–98)	97 (95–98)	0 (0–1.8)	0 (0–1.8)	1.5 (0.5–4.2)	5.2 (3.2–8.3)	5.2 (3.2–8.3)	3.8 (2.2–6.6)
CFZ	79 (75–82)	64 (59–68)	87 (83–89)	2.2 (1.0–4.8)	6.4 (4.0–10)	24 (19–29)	41 (35–48)	69 (63–74)	2.0 (0.9–4.6)
CFE	73 (67–77)	77 (72–82)	92 (88–95)	22 (17–28)	4.8 (2.6–8.6)	10 (7.0–15)	40 (31–49)	54 (45–63)	3.0 (1.0–8.4)
CEPH	88 (86–89)	84 (82–86)	92 (90–93)	6.1 (4.7–7.8)	3.0 (2.1–4.4)	13 (11–16)	19 (16–21)	27 (25–30)	3.0 (2.0–4.6)
MER	88 (85–91)	89 (86–92)	100 (99–100)	12 (9.4–15)	11 (8.1–14)	0 (0–0.9)	^c	^c	^c
IMI	NA ^b	NA ^b	100 (99–100)	NA ^b	NA ^b	0.2 (0–1.3)	NA ^b	NA ^b	100 (5.5–100)
CARB	88 (84–90)	89 (86–92)	100 (99–100)	12 (9.4–15)	11 (8.1–14)	0 (0–0.6)	^c	^c	100 (5.5–100)
CIP (FQ)	92 (88–94)	92 (88–94)	96 (94–97)	0 (0–1.4)	0 (0–1.4)	8.3 (5.5–12)	18 (14–23)	18 (14–23)	0 (0–1.5)
GEN	96 (94–98)	96 (94–98)	98 (96–99)	0.5 (0.1–1.7)	0.5 (0.1–1.7)	0.5 (0.1–1.7)	17 (11–25)	17 (11–25)	9.7 (1.7–15)
TOB	97 (95–98)	97 (95–98)	98 (96–99)	1.3 (0.6–3.1)	1.3 (0.6–3.1)	1.9 (0.9–3.8)	8.3 (4.6–15)	8.3 (4.6–15)	4.2 (1.8–9.4)
AMIN	97 (95–98)	97 (95–98)	98 (96–98)	0.9 (0.4–1.8)	0.9 (0.4–1.8)	1.1 (0.6–2.2)	12 (8.5–17)	12 (8.5–17)	6.7 (4.1–11)
COT	NA ^b	NA ^b	97 (94–98)	NA ^b	NA ^b	2.2 (1.0–4.6)	NA ^b	NA ^b	5.1 (2.9–8.6)
COL	NA ^b	NA ^b	99 (96–99)	NA ^b	NA ^b	1.4 (0.6–3.6)	NA ^b	NA ^b	100 (20–100)
MISC	NA ^b	NA ^b	97 (96–98)	NA ^b	NA ^b	1.5 (0.8–2.8)	NA ^b	NA ^b	5.9 (3.5–9.6)
ALL ^d	87 (86–88)	86 (85–87)	93 (92–94)	5.1 (4.4–5.9)	4.3 (3.6–5.0)	5.1 (4.5–5.8)	26 (24–28)	29 (27–31)	11 (9.2–12)

The data consisted of a combination of two datasets: a random sample of *E. coli* bacteraemia (*n* = 234) (Appendix S6), and third-generation cephalosporin-resistant *E. coli* UTI and bacteraemia (*n* = 283) (Appendix S6).

^a Antimicrobial: AMO, amoxicillin; SSP, small-spectrum penicillin (AMO); AMC, amoxicillin-clavulanic acid; PITA, piperacillin-tazobactam; BRSP, broad-spectrum penicillins (sum of AMO and PITA); CER, cefuroxime; CFO, cefotaxime; CFZ, ceftazidime; CFE, cefepime; CEPH, cephalosporins (sum of CER, CFO, CFZ, and CFE); MER, meropenem; IMI, imipenem; CARB, carbapenems (sum of MER, and IMI); CIP, ciprofloxacin; FQ, fluoroquinolones (CIP); GEN, gentamicin; TOB, tobramycin; AMIN, aminoglycosides (sum of GEN and TOB); COT, cotrimoxazole; COL, colistin; MISC, miscellaneous agents (sum of COT and COL).

^b Antimicrobial was not available in the phenotype prediction tool.

^c Output not calculated because of the absence of phenotypic resistant isolates in dataset.

^d Calculated with the sum of all observed true positives, true negatives, false positives, and false negatives.

amoxicillin meeting both criteria (MER: 1.6% (0.5–5.8), VMER: 1.8% (0.9–3.6)) (Table 1).

Overall MERs for KOVER-AMR-SCM, KOVER-AMR-CART, and ResFinder 4.1 were 5.1% (4.4–5.9), 4.3% (3.6–5.0), and 5.1% (4.5–5.8), respectively. However, for KOVER-SCM 6/11 of the assessed antimicrobials, for KOVER-CART 5/11, and 9/13 of the assessed antimicrobials of ResFinder 4.1 predictions MERs were below 3% (FDA criteria), indicating acceptable MERs for the majority of assessed antimicrobials (Table 1). Observed MERs were particularly high for cefepime (4.8–22%) and meropenem (11–12%) for KOVER-AMR-models and for piperacillin-tazobactam 16% 95% CI (13–20%) and ceftazidime 24% (19–29%) for ResFinder 4.1 (Table 1).

KOVER-AMR-SCM, KOVER-AMR-CART, and ResFinder 4.1 VMERs for all antimicrobials were 26% (24–26), 29% (27–31), and 11% (9.2–12), exceeding the FDA criteria. KOVER-AMR-models did not achieve VMERs below FDA criteria for any of the assessed antimicrobials, while for ResFinder 4.1, VMERs for 4/13 antimicrobials were within the limits of the FDA criteria, namely amoxicillin, cefotaxime, ceftazidime, and ciprofloxacin. For all models, observed VMERs were particularly high for amoxicillin-clavulanic acid (47–57%) and piperacillin-tazobactam (39–88%) (Table 1).

Overall observed concordance was 87% (86–88) for KOVER-AMR-SCM and 86% (85–87) for KOVER-AMR-CART, indicating no preference for either of KOVER-AMR-models. Overall concordance of ResFinder 4.1 was 93% (92–94), confirming the better performance of this tool. Furthermore, for both KOVER-AMR-models and ResFinder 4.1 concordances were high for aminoglycosides (97–98%), while concordances for broad-spectrum penicillins were low (76–79%), indicating differences per antimicrobial class in phenotype predictions (Table 1).

Lastly, comparing results from the random sample of *E. coli* and 3rdGCR-*E. coli* showed a higher concordance for the first: KOVER-AMR-SCM: 92% versus 82%, KOVER-AMR-CART: 93% versus 80%, ResFinder 4.1: 96% versus 91% (Appendix S6). Furthermore, overall MERs were lower in the random sample, while VMERs were higher, compared to the 3rdGCR-*E. coli* (Appendix S6). Indicating difference in tool performance in different datasets with amongst others different resistance rates.

Assessment of discrepancies

In total, 727 discrepancies in genotype-to-phenotype predictions were observed for KOVER-AMR-SCM, 773 for KOVER-AMR-CART, and 424 for ResFinder 4.1 (Table 2). Grouping these into possible explanatory categories showed that 36% of all discrepancies occurred in isolates with a MIC with a factor 2 above or below the clinical breakpoint. Presence of a genetic predictor that did not predict phenotypic resistance occurred evenly between the

tools, while both KOVER-AMR-models were more likely to miss a genetic predictor for a resistant phenotype compared to ResFinder 4.1 (Table 2). For amoxicillin-clavulanic acid often all tools failed to predict resistance. For piperacillin-tazobactam 261 discrepancies were observed; MERs for piperacillin-tazobactam occurred more often in ResFinder 4.1, where presence of the *bla*_{OXA-1} gene often did not translate into phenotypic resistance against this penicillin/inhibitor combination. VMERs for piperacillin-tazobactam occurred more often in both KOVER-AMR tools. Lastly, ceftazidime, cefepime, and meropenem models of KOVER-AMR-SCM and/or KOVER-AMR-CART contained genetic predictors for resistance that did not result in phenotypic resistance in our strain sets.

Discussion

This study externally validated two WGS-AST tools, KOVER-AMR-models and ResFinder 4.1, that provide direct phenotype predictions, in 517 *E. coli* isolates recovered from patients with infections in the Netherlands. Overall, ResFinder 4.1 performed better than KOVER-AMR-models; however, none of the tools achieved overall (V)MERs below FDA criteria for the total assessed panel of antimicrobials. MERs ≤3% were achieved for 6 and 5/11 tested antimicrobials for KOVER-AMR-models and for 9/13 antimicrobials tested with ResFinder 4.1. VMERs with a 95% CI ≤ 1.5 to ≤7.5 was achieved in none of the tested antimicrobials for KOVER-AMR and for 4/13 tested antimicrobials with ResFinder 4.1. Only phenotype predictions for amoxicillin resistance by ResFinder 4.1 achieved both FDA criteria.

We compared the observed (V)MERs and concordances with the reported internal validations of these tools (Appendix S7). The phenotypic resistance rates of the *E. coli* dataset used for internal validation of KOVER-AMR-models were comparable to this study, while phenotypic resistance percentages of the datasets used for ResFinder 4.1 were slightly higher. The reported overall MERs and VMERs of KOVER-AMR-models and ResFinder 4.1 were lower than our observations [7,17]. The observed differences between the reported internal validations of these tools and our results can have several explanations. Firstly, KOVER-AMR-models were created by supervised machine learning algorithm KOVER, where KOVER selected *k*-mers that were able to predict resistance to a certain antimicrobial. By default, no underlying biological mechanism (i.e. causal relation) is needed for selection of predictors. The risk of this methodology is that the model might not predict resistance in a new dataset, if the dataset used for model training does not represent the population in which the model will be applied. We observed this for the models predicting resistance against cefepime and meropenem. Furthermore, if the training dataset of KOVER-AMR-models had an absence or too low prevalence of certain

Table 2
Summarization of observed discrepancies ((V)MEs) in possible categories

Discrepancy category	KOVER-AMR-SCM % (n)	KOVER-AMR-CART % (n)	ResFinder 4.1 % (n)
MIC around clinical breakpoint ^a	36% (262)	35% (269)	36% (154)
likely incorrect genetic predictor ^b	20% (145)	16% (125)	33% (139)
possible incorrect genetic predictor or incorrect phenotype ^c	1% (9)	1% (9)	4% (18)
likely missing genetic predictor ^d	32% (237)	38% (296)	6% (26)
possible missing genetic predictor or incorrect phenotype ^e	10% (74)	10% (74)	21% (87)
Total	100% (727)	100% (773)	100% (424)

^a MIC around clinical breakpoint: minimum inhibitory concentration was ≤2x above/below the clinical breakpoint [21].

^b Likely incorrect genetic predictor (ME) if ≥ 1 of the assessed tools predicted the isolate as susceptible.

^c Possible incorrect genetic predictor or incorrect phenotype (ME) if all of the assessed tools predicted the isolate as resistant.

^d Likely missing genetic predictor (VME) if ≥ 1 of the assessed tools predicted the isolate as resistant.

^e Possible missing genetic predictor or incorrect phenotype (VME) if all of the assessed tools predicted the isolate as susceptible.

resistance genes/mutations, these will not be included as genetic resistance predictors [14]. For example, we observed this in the models predicting amoxicillin resistance, where *bla*_{CTX-M} genes were missed. Both issues can be overcome when the training datasets used for KOVER are large and diverse enough, which will result in generalizability to a real-life setting [14]. The great advantage of machine learning tools like KOVER is that they can be used to discover new resistance traits, a feature that is absent from reference databases like ResFinder. Secondly, Bortolaia et al. [7] did not include amoxicillin-clavulanic acid or piperacillin-tazobactam, for which we observed the highest (V)MERs within the assessed antimicrobials for ResFinder 4.1 predictions. For ResFinder 4.1, the most important resistance predictor for piperacillin-tazobactam was the *bla*_{OXA-1} gene [23–25]. However, in our study the presence/absence of *bla*_{OXA-1} often did not predict the phenotype correctly. Possible explanations for false negatives/positives may be: (1) altered activity of efflux-pumps, (2) altered membrane permeability, (3) gene-duplication, (4) increased expression of chromosomal AmpC or ESBL-gene, or (5) non-expression of *bla*_{OXA-1} [24,26]. Reference databases like ResFinder 4.1 will most likely not be able to detect these alterations, as they are not captured in gene presence/absence. Supervised machine learning algorithms might be able to detect these characteristics. For example, if these factors would occur more often in certain subpopulations, an association of piperacillin-tazobactam resistance with the population structure could be detected [26,27].

To our knowledge this is the first study that externally validated direct phenotype prediction tools: KOVER-AMR-models and ResFinder 4.1 for *E. coli* isolates recovered from clinical practice. We additionally compared our results to the FDA criteria needed for diagnostic use. Both are needed for progress to future implementation of WGS-AST tools in clinical practice [12,28]. Furthermore, we evaluated the tools in a random sample of *E. coli* blood isolates, representing a diverse and representative sample of Dutch clinical practice. We also included 3rdGCR-*E. coli* isolates; the higher proportion of phenotypic resistance allowed us to more accurately calculate VMERs, arguably the most relevant outcome measure for patient care.

This external validation study has several potential limitations. Firstly, we did not have the possibility to re-analyse discrepant phenotypes by repeating culture-AST, which could have contributed to the observed high (V)MERs. Gordon et al. mentioned that in their study 40% of the discrepancies were solved by re-analysing discrepant phenotypes [29]. Indeed, 36% of the discrepancies in this study occurred in isolates with an MIC with a factor 2 above or below the clinical breakpoint. Secondly, we did not use the gold-standard for culture-AST, namely broth microdilution [3]. Instead, we used VITEK 2, and BD Phoenix, two frequently used methodologies in clinical practice, which may have influenced the observed MICs. Thirdly, the sampling from a confined area in time and space (i.e. the Netherlands between 2014 to 2017) might have influenced the results. Lastly, we used the default minimal gene length in ResFinder 4.1 of $\geq 60\%$. However, a minimal gene length of for example 80% would have resulted in six fewer MEs and nine additional VMEs (Appendix S4). As VMEs have the potential to lead to adverse patient outcomes because of undertreatment, the cut-off of 60% was justifiable in this dataset. However, the optimal cut-off might differ per setting.

Although neither of the tools fulfil the criteria to be used for clinical decision making, the tools did provide concordant results in 86% to 87% (KOVER-AMR) and 93% (ResFinder 4.1) of isolates when compared with culture-AST. This is in line with observed concordances in proof of principle WGS-AST studies [14–16]. These observations indicate that there are likely enough possibilities for improvements to meet FDA criteria. For this the following route can be proposed. Firstly, collective effort should be put in making phenotype-genotype datasets available [14]. These datasets should

be used for a continued search for new resistance traits with machine learning algorithms like, for example, KOVER, or recently developed methodologies like pyseer or INGOT-DR [17,27,30]. The resulting output should be used to update a centrally curated and freely available reference database like ResFinder 4.1. Secondly, the scientific community should focus on feasible and affordable direct sample sequencing applications for clinical settings, starting with clinical samples that are normally sterile like blood [9–11]. Thirdly, more external validation studies for different pathogens and datasets are needed, as well as clinical implementation studies [14]. This effort should be done as it has the potential to improve the quality and speed of obtaining information regarding the optimal treatment of infections.

Transparency declaration

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. All authors have nothing to declare. All generated raw reads were submitted to the European Nucleotide Archive (ENA) of the European Bioinformatics Institute (EBI) under the study accession number PRJEB35000. All phenotypes are available in the supplementary material (Supplementary material 2). This study was internally funded by the Netherlands National Institute for Public Health and the Environment (RIVM) and the University Medical Center Utrecht (UMCU) and the authors received no specific funding for this work.

Author contributions

All authors contributed to the study design or conduct. TV, TB, VM, RW, and JK wrote the study protocol. TB, RW, and JK obtained funding. TV and VM performed data collection. TV performed bioinformatic and statistical analysis. TB, RW, and JK supervised the study. TV drafted the manuscript. All authors reviewed the manuscript throughout the writing process.

Acknowledgements

We would sincerely like to thank Kim van der Zwaluw, Carlo Verhulst, and Judith Vlooswijk for their contributions in the laboratory execution of the study. We also would sincerely like to thank Denise van Hout, Marc Bonten, and Ascelijn Reuland for their contributions in data collection. Finally, we would like to thank Anita Schürch in her contribution to the assembly of the sequences.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cmi.2022.05.024>.

References

- [1] European Centre for Disease Prevention and Control. Surveillance of antimicrobial resistance in Europe 2018. Stockholm: ECDC; 2019. Available from: <https://www.ecdc.europa.eu/sites/default/files/documents/surveillance-antimicrobial-resistance-Europe-2018.pdf>.
- [2] Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect Dis* 2019;1:56–66. [https://doi.org/10.1016/S1473-3099\(18\)30605-4](https://doi.org/10.1016/S1473-3099(18)30605-4).
- [3] Jorgensen JH, Ferraro MJ. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. *Clin Infect Dis* 2009;49: 1749–55. <https://doi.org/10.1086/647952>.
- [4] Török ME, Peacock SJ. Rapid whole-genome sequencing of bacterial pathogens in the clinical microbiology laboratory—pipe dream or reality? *J Antimicrob Chemother* 2012;67:2307–8. <https://doi.org/10.1093/jac/dks247>.
- [5] Deurenberg RH, Bathoorn E, Monika A, Couto N, Ferdousa M, García-Cobos S, et al. Application of next generation sequencing in clinical microbiology and

- infection prevention. *J Biotechnol* 2017;243:16–24. <https://doi.org/10.1016/j.jbiotec.2016.12.022>.
- [6] Andersson DI, Balaban NQ, Baquero F, Courvalin P, Glaser P, Gophna U, et al. Antibiotic resistance: turning evolutionary principles into clinical reality. *FEMS Microbiol Rev* 2020;2:171–88. <https://doi.org/10.1093/femsre/fuaa001>.
 - [7] Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020;12:3491–500. <https://doi.org/10.1093/jac/dkaa345>.
 - [8] Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet* 2019;6:341–55. <https://doi.org/10.1038/s41576-019-0113-7>.
 - [9] Su M, Satola SW, Read TD. Genome-based prediction of bacterial antibiotic resistance. *J Clin Microbiol* 2019;3:e01405–18. <https://doi.org/10.1128/JCM.01405-18>.
 - [10] Takt AM, Avershina E, Frye SA, Naseer U, Ahmad R. Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. *Sci Rep* 2020;1:7622. <https://doi.org/10.1038/s41598-020-64616-x>.
 - [11] Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016;7589:228–32. <https://doi.org/10.1038/nature16996>.
 - [12] Peiffer-Smadja N, Dellièvre S, Rodriguez C, Birgand G, Lescure FX, Fourati S, et al. Machine learning in the clinical microbiology laboratory: has the time come for routine practice? *Clin Microbiol Infect* 2020;10:1300–9. <https://doi.org/10.1016/j.cmi.2020.02.006>.
 - [13] U.S. Department. Of health and human services food and Drug administration. Guidance for industry and FDA. Class II special controls guidance document: antimicrobial susceptibility test (AST) systems. Rockville: FDA; 2009. Available from: <https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/antimicrobial-susceptibility-test-ast-systems-class-ii-special-controls-guidance-industry-and-fda>.
 - [14] Hendriksen RS, Bortolaia V, Tate H, Tyson GH, Aarestrup FM, McDermott PF. Using genomics to track global antimicrobial resistance. *Front Public Health* 2019;7:242. <https://doi.org/10.3389/fpubh.2019.00242>.
 - [15] Moradigaravand D, Palm M, Farewell A, Mustonen V, Warringer J, Parts L. Prediction of antibiotic resistance in *Escherichia coli* from large-scale pan-genome data. *PLOS Comput Biol* 2018;12:1–17. <https://doi.org/10.1371/journal.pcbi.1006258>.
 - [16] Mahfouz N, Ferreira I, Beisken S, von Haeseler A, Posch AE. Large-scale assessment of antimicrobial resistance marker databases for genetic phenotype prediction: a systematic review. *J Antimicrob Chemother* 2020;75:3099–108. <https://doi.org/10.1093/jac/dkaa257>.
 - [17] Drouin A, Letarte G, Raymond F, Marchand M, Corbeil J, Laviolette F. Interpretable genotype-to-phenotype classifiers with performance guarantees. *Sci Rep* 2019;9:1–13. <https://doi.org/10.1038/s41598-019-40561-2>.
 - [18] van Hout D, Verschuuren TD, Bruijning-Verhagen PCJ, Bosch T, Schürch AC, Willems RJJ, et al. Extended-spectrum beta-lactamase (ESBL)-producing and non-ESBL-producing *Escherichia coli* isolates causing bacteremia in The Netherlands (2014–2016) differ in clonal distribution, antimicrobial resistance gene and virulence gene content. *PLoS One* 2020;15:e0227604. <https://doi.org/10.1371/journal.pone.0227604>.
 - [19] Verschuuren TD, van Hout D, Arredondo-Alonso S, Fluit AC, Reuland EA. Top J et al. Comparative genomics of ESBL-producing *Escherichia coli* (ESBL-Ec) reveals a similar distribution of the 10 most prevalent ESBL-Ec clones and ESBL genes among human community faecal and extra-intestinal infection isolates in The Netherlands (2014–17). *J Antimicrob Chem* 2021;4:901–8. <https://doi.org/10.1093/jac/dkaa534>.
 - [20] Commissie Regelgeving Onderzoek (COREON). Gedragscode Gezondheidsonderzoek. 2022. Available from: <https://www.coreon.org/wp-content/uploads/2022/01/Gedragscode-Gezondheidsonderzoek-2022.pdf>. [Accessed 29 August 2022].
 - [21] The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters Version 9.0. Växjö. EUCAST; 2019. Available from: https://eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_9.0_Breakpoint_Tables.pdf.
 - [22] Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, et al. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res* 2020;D1:D606–12. <https://doi.org/10.1093/nar/gkz943>.
 - [23] Sugumar M, Kumar KM, Manoharan A, Anbarasu A, Ramaiah S. Detection of OXA-1 β -lactamase gene of *Klebsiella pneumoniae* from blood stream infections (BSI) by conventional PCR and in-silico analysis to understand the mechanism of OXA mediated resistance. *PLoS One* 2014;3:e91800. <https://doi.org/10.1371/journal.pone.0091800>.
 - [24] Livermore DM, Day M, Cleary P, Hopkins KL, Toleman MA, Wareham DW, et al. OXA-1 β -lactamase and non-susceptibility to penicillin/ β -lactamase inhibitor combinations among ESBL-producing *Escherichia coli*. *J Antimicrob Chemother* 2019;2:326–33. <https://doi.org/10.1093/jac/dky453>.
 - [25] Henderson A, Paterson DL, Chatfield MD, Tambyah PA, Lye DC, De PP, et al. Association between minimum inhibitory concentration, beta-lactamase genes and mortality for patients treated with piperacillin/tazobactam or meropenem from the MERINO study. *Clin Infect Dis* 2021;11:e3842–50. <https://doi.org/10.1093/cid/ciaa1479>.
 - [26] Kaldalu N, Hauryliuk V, Tenson T. Persisters—as elusive as ever. *Appl Microbiol Biotechnol* 2016;100:6545–53. <https://doi.org/10.1007/s00253-016-7648-8>.
 - [27] Lees JA, Tien Mai T, Galardini M, Horsfield ST, Parkhill J, Corander J. Improved prediction of bacterial genotype-phenotype associations using interpretable pan-genome-spanning regressions. *mBio* 2020;4. <https://doi.org/10.1128/mBio.01344-20>. e01344-20.
 - [28] Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;61:1085–94. <https://doi.org/10.1016/j.jclinepi.2008.04.008>.
 - [29] Gordon NC, Price JR, Cole K, Everitt R, Morgan M, Finney J, et al. Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *J Clin Microbiol* 2014;52:1182–91. <https://doi.org/10.1128/JCM.03117-13>.
 - [30] Zabati Z, Dexter N, Amir Safari AH, Libbrecht M, Chindelevitch L, INGOT-DR: an interpretable classifier for predicting drug resistance in *M. tuberculosis* Algorithms Mol Biol 2021;1:17. <https://doi.org/10.1186/s13015-021-00198-1>.