AJCP | ORIGINAL ARTICLE

Cycle Threshold Probability Score for Immediate and Sensitive Detection of B.1.351 SARS-CoV-2 Lineage

Dieter De Smet, MD,[1] *Merijn Vanhee, MSc, PhD,*[1] *Brigitte Maes, MD, PhD,*[2] *Koen Swaerts, MSc,*[1] *Peter De Jaeger, MScEng, PhD,*[1] *Karel Maelegheer, MD,*[3] *Frederik Van Hoecke, MD,*[1] *and Geert Antoine Martens, MD, PhD*[1,4]

From the [1]Department of Laboratory Medicine, AZ Delta General Hospital, Roeselare, Belgium; [2]Department of Clinical Biology, Jessa Hospital, Hasselt, Belgium; [3]Department of Clinical Biology, AZ Sint-Lucas Hospital, Bruges, Belgium; and [4]Department of Biomolecular Medicine, Ghent University, Gent, Belgium.

ABSTRACT

Background: Detection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants of concern associated with immune escape is important to safeguard vaccination efficacy. We describe the potential of delayed *N* gene amplification in the Allplex SARS-CoV-2 Assay (Seegene) for screening of the B.1.351 (20H/501.V2, variant of concern 2 [VOC.V2], South African SARS-CoV-2 variant) lineage.

Methods: In a study cohort of 397 consecutive polymerase chain reaction–positive samples genotyped by whole-genome sequencing, amplification curves of *E/N/S-RdRP* targets indicated delayed *N* vs *E* gene amplification characteristic of B.1.351. Logistic regression was used to calculate a VOC.V2 probability score that was evaluated as a separate screening test in an independent validation cohort vs sequencing.

Results: B.1.351 showed a proportionally delayed amplification of the *N* vs *E* gene. In logistic regression, only *N* and *E* gene cycle thresholds independently contributed to B.1.351 prediction, allowing calculation of a VOC.V2 probability score with an area under the curve of 0.94. At an optimal dichotomous cutoff point of 0.12, the VOC.V2 probability score achieved 98.7% sensitivity at 79.9% specificity, resulting in a negative predictive value (NPV) of 99.6% and a positive predictive value of 54.6%. The probability of B.1.351 increased with an increasing VOC.V2 probability score, achieving a likelihood ratio of 12.01 above 0.5. A near-maximal NPV was confirmed in 153 consecutive validation samples.

Conclusions: Delayed *N* vs *E* gene amplification in the Allplex SARS-CoV-2 Assay can be used for fast and highly sensitive screening of B.1.351.

KEY POINTS

- The high mutation rate of coronaviruses results in shifts of cycle threshold (Ct) values in multigene-target in vitro diagnostic medical device regulation polymerase chain reaction (PCR) assays.
- We provide a theoretical framework to build simple logistic regression models based on Ct values of a widely used commercial severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) PCR assay.
- This framework is validated against whole-genome sequencing as a gold standard for screening the B.1.351 variant of concern using the Allplex SARS-CoV-2 Assay.

KEY WORDS

PCR; Cycle threshold value; Modeling; Data analysis; SARS-CoV-2; Secondary data analysis

Am J Clin Pathol May 2022;157:731-741
HTTPS://DOI.ORG/10.1093/AJCP/AQAB186

Received: August 6, 2021
Accepted: September 21, 2021
Advance publication: November 1, 2021

Corresponding author: Geert A. Martens, MD; geert.martens@azdelta.be.

Funding: This work was supported by a private donation by board members of Fagron (Nazareth, Belgium), a health care company, to RADar, the teaching and education initiative of AZ Delta General Hospital, to be used as an unconditional research grant for data collection, collaborative collaboration, and open access publication. The sponsor had no influence on the study design, data interpretation, and drafting of the manuscript.

INTRODUCTION

In December 2020, an unexpected rise in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in the United Kingdom was attributed to the emergence of a new SARS-CoV-2 variant of concern (VOC), lineage B.1.1.7 (20I/501.V1, VOC 202012/01), first detected by an efficient national-scale genomic surveillance network. B.1.1.7 shows a 40% to 70% higher transmission rate than the ancestral wild-type SARS-CoV-2 lineages first identified in Wuhan (A/B.1),[1,2] mainly explained by the N501Y amino acid substitution in the spike-receptor

© The Author(s) 2021. Published by Oxford University Press on behalf of American Society for Clinical Pathology. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/ajcp/article/157/5/731/6415253 by Hasselt University user on 01 June 2022

binding domain (S-RBD) that increases the affinity of S-RBD for the human angiotensin-converting enzyme 2 receptor 7- to 19-fold.[3,4] B.1.1.7 rapidly became the dominant strain in the United Kingdom and many European countries.[5] Its spread could be accurately monitored by an accidental interference by 1 of its defining mutations, the spike 69/70 deletion (69.70del), in the widely used TaqPath polymerase chain reaction (PCR) (Thermo Fisher Scientific), resulting in a characteristic signature with preserved amplification of *N* gene and *ORF1ab* gene targets but *S* gene target failure (SGTF, or *S* dropout). Though recent data indicate that B.1.1.7 is overall 60% more deadly than the ancestral B.1 lineage,[6,7] it does not appear to display immune escape. B.1.1.7 appears refractory to N-terminal domain (NTD)–targeting antibodies but not to the S-RBD–neutralizing antibodies induced by all currently approved vaccines,[4,8] with generally preserved in vitro neutralization convalescent sera after wild-type virus infection and by immune sera after the BNT162b2 (Pfizer/BioNTech) and AZD1222 (AstraZeneca/Oxford)[4,9,10] vaccines.

In an example of convergent evolution, the N501Y mutation independently arose in several other SARS-CoV-2 lineages, such as the B.1.351 (20H/501.V2, VOC 2) in South Africa[11] and the P.1 (20J/501Y.V3, VOC 3) in Brazil.[12] Both lineages had higher infectivity than the wild-type virus and rapidly achieved regional dominance. In another, more concerning example of convergent evolution, the B.1.351 and P.1 lineages share the additional E484K spike mutation that was independently confirmed as a powerful driver of clinical immune escape. This partial immune escape explains the resurgence of coronavirus disease 2019 (COVID-19) in Manaus, Brazil, by P.1 in a population with 76% seroprevalence[13] and the absence of protective immunity of prior wild-type infection to B.1.351 in the placebo arms of the NVX-CoV2373 vaccine trial in South Africa.

The impact of various SARS-CoV-2 mutations/variants on immune escape was mainly derived from several small-scale in vitro pseudovirus neutralization studies using polyclonal convalescent sera or monoclonal antibodies after wild-type (non-VOC) infections and polyclonal sera after vaccination with the current first-generation S-RBD–targeted vaccines. Introduction of E484K in a B.1.1.7 background leads to a 6-fold lower neutralization by convalescent sera.[3,4] Also, vaccine efficacy is reduced: P.1 shows a moderately (2- to 3-fold) lower neutralization by messenger RNA-1273 and BNT162b2 vaccine sera.[8] In the case of B.1.351, the effect is more pronounced, likely because of the additional effects of the NTD mutations in this lineage, leading to 6- to 12-fold lower neutralization by both vaccines.[8,14,15] These effects are consistently confirmed in vivo. Recent data collected in South Africa showed a drop in vaccine efficacy to prevent moderate to severe COVID-19 by the B.1.351 lineage compared with non-VOC strains: from 95.6% to 50% for the NVX-CoV2373 (Novavax) vaccine, from 72% to 57% for Ad26.COV2.S (Janssen), and a complete loss of efficacy of the AZD1222 vaccine (66.7% for all variants to 10.6% for B.1.351).[16]

These findings indicate that early detection of emerging variants with immune escape, such as B.1.351 and P.1, at the population level is important for a data driven monitoring of vaccination efficacy, even though the impact on individual patient management is minimal. Intense population-wide surveillance requires a combination of several test strategies: unbiased surveillance by whole-genome sequencing (WGS), the flexible introduction of targeted reflex PCR assays for lineage-defining mutations in the spike protein (ie, at positions N501, D253, or Q677) or mutations associated with potential immune escape (ie, at positions E484, L452, or S477) depending on regional strain prevalence but also the intelligent use of subtle variations in standard SARS-CoV-2 PCR amplification curves. The latter could allow fast, high-throughput, low-cost screening of emerging VOCs, as illustrated by the impact of the SGFT/*S* gene dropout for B.1.1.7 surveillance,[1,17,18] rendering variant screening available for health care systems with limited resources.

In this article we describe delayed amplification of the *N* gene target characteristic of the B.1.351 variant in the Allplex SARS-CoV-2 Assay (Seegene), which is currently used in high volumes in more than 70 countries. The *N* gene delay was investigated in a study cohort of SARS-CoV-2 isolates containing a representative number of wild-type lineages B.1.1.7 and B.1.351. A probability score was calculated for the presence of B.1.351 VOC 2 (VOC.V2 probability score) based on *E/N/S-RdRP* cycle threshold (Ct) values. We subsequently studied the diagnostic power of this new VOC.V2 probability score as separate test result in an independent validation cohort, as a stand-alone test, and in conjunction with N501Y and 69.70del mutation–specific PCR assays.

MATERIALS AND METHODS

Samples

The study was performed on consecutive samples of subjects (inpatients and outpatients) who underwent SARS-CoV-2 PCR testing and WGS as part of routine diagnostic workup and baseline or active surveillance for VOCs; samples were allocated to a study cohort or a validation cohort. The study cohort consisted of 397 unique viral isolates, including 141 501Y.V1 clades (UK variant—hence, the B.1.1.7 lineage), 78 501Y.V2 clades (South African variant—hence, the B.1.351 lineage), and 178 samples with other SARS-CoV-2 clades that did not belong to any of the 3 currently recognized VOC clades (hence, non-VOC lineages). The study cohort was representative of regional prevalence as measured by unbiased surveillance by the Belgian National WGS consortium for which AZ Delta acts as a sentinel hub. Non-VOC Pangolin lineages included B.1.0, B.1.162, B.1.214, B.1.221, B.1.258, B.1.160, B.1.177, B.1.1.222, and B.1.1.29. The validation cohort contained 153 unique viral isolates: 60 non-VOCs, 57 B.1.1.7 clades, 29 B.1.351 clades, and 7 20J/501Y.V3 clades (Brazilian variant—hence, P.1). For detailed lineage composition of study and validation cohorts, see Supplemental Table 1 (all supplemental materials can be found at *American Journal of Clinical Pathology* online). The study was performed on analytical leftover samples and secondary use of data and samples obtained during standard-of-care diagnostic workup. It was performed according to the Helsinki Declaration, with full respect for individuals' right to confidentiality and with a waiver of informed consent for COVID-19–related diagnostics and modeling, approved by the AZ Delta ethical committee (Clinical Trial No. IRB B1172020000008).

Downloaded from https://academic.oup.com/ajcp/article/157/5/731/6415253 by Hasselt University user on 01 June 2022

© American Society for Clinical Pathology

RNA Extraction, Reverse Transcription–PCR for Detection of SARS-CoV-2, and Reflex PCR for *N501Y* and 69.70del

RNA was extracted from nasopharyngeal swabs using the STARMag 96 × 4 Viral DNA/RNA 200 C Kit (Seegene Technologies) on the Hamilton STARlet workstation, followed by real-time PCR using the Allplex SARS-CoV-2 Assay. PCR amplification was run on a CFX96 real-time thermal cycler (Bio-Rad Laboratories), and data were analyzed with the SARS-CoV-2 Viewer (Seegene). The assay simultaneously detects 4 different SARS-CoV-2 genes, resulting in separate Ct values for the *E* and *N* genes and 1 combined Ct value for the *RdRP* and *S* genes (*S/RdRP*) (for the raw data, see Supplemental Table 1). For a subset of PCR-positive samples in the study and validation cohorts (for the raw data, see Supplemental Table 2), 2 key mutations for VOCs in the spike protein were detected by reflex testing using the Novaplex SARS-CoV-2/H-UK (N501Y and 69-70del) Assay (Seegene). Therefore, a new RNA extraction was performed on the original sample, using the STARMag 96 × 4 Universal kit, followed by manual PCR setup and amplification on a CFX96 thermal cycler. Reflex PCR targets the N501Y amino acid substitution (A23063T) characteristic of VOC.V1-V3 and the HV 69-70 deletion (21,765-21,770 deletion—hence, 69.70del) that, in the presence of the N501Y mutation, is highly suggestive of B.1.1.7. A conserved sequence in the *RdRP* gene is co-amplified as internal control, and negative results for N501Y/69.70del were considered true negative only if the *RdRP* internal control had a Ct value less than 33. For multicenter interlaboratory validation, a secondary data analysis was performed on a second validation set of n = 308 consecutive samples analyzed by an independent laboratory (AZ Sint-Lucas Brugge Hospital) with the Allplex SARS-CoV-2 Assay using an alternative RNA extraction on the Maelstrom 9600 platform (Taiwan Advanced Nanotech [TANBead]) followed by thermal cycling on the CFX96 device.

SARS-CoV-2 WGS and Clade Calling

As the reference method for genomic fingerprinting of viral clades, WGS was performed using the research-use-only AmpliSeq for Illumina SARS-CoV-2 Research Panel on the Illumina MiSeq sequencer (40 samples on MicroV2 flow cell) according to the manufacturer's standard protocol: 7 µL RNA was reverse-transcribed using AmpliSeq complementary DNA synthesis for Illumina, followed by amplification of 237 virus-specific amplicons covering more than 99% of the 30-kb reference genome, aiming at a median coverage above 200×, minimal coverage for mutation calling of 10×, a minimum of 30,000 reads per sample, and a maximum of 1-kb bases below minimal coverage. A consensus sequence was constructed using an in-house pipeline containing Trimmomatic for trimming (USADELLAB.org), alignment by the Burrows-Wheeler Aligner, mutation calling by Freebayes, and inspection of sequence quality using the Integrative Genomics Viewer. Clade calling on the consensus FASTA file was done by both Pangolin lineage assignment (version 2.3.2, https://pangolin.cog-uk.io) and Nextclade (version 0.14.1, https://clades.nextstrain.org) web tools, and only concordant clade calling results were used for further analysis (Supplemental Table 1). All sequences were uploaded to GISAID.

Statistics

Statistical analysis was performed using MedCalc software, version 12.2.1. Data are expressed as medians (25th-75th percentiles), and we used the Mann-Whitney *U* test to test statistical difference between groups. Confidence intervals (CIs) for likelihood ratios (LRs) were calculated using the method described by Simel et al.[7] Comparisons of Ct values within and between sample subsets were made using Spearman rank correlation and Passing-Bablok regression analysis. Concordance analysis of integrated results for N501Y/69.70del reflex PCR and VOC.V2 probability scores with WGS was conducted by calculation of Cohen κ agreement coefficient (linear weights).

Different prediction models were trained using the Python scikit-learn library (ie, logistic regression, support vector machines, random forest trees, BaggingClassifier, and gaussian Naive bayes). Best accuracy and area under the curve (AUC) was achieved using logistic regression (for an overview of accuracy and the AUC of various tested models, see Supplemental Table 3). The logistic regression classifier was validated and further adopted in the MedCalc statistical tool, which offers the benefit of interpretability. The cutoff for the logistic regression model was optimized to yield optimal recall for detecting the presence of the B.1.351 variant.

RESULTS

Distinct *E, N,* and *S/RdRP* Ct Profiles Among SARS-CoV-2 Viral Lineages

Retrospective analysis of the study cohort (n = 397) indicated variations in the distribution profiles of *E/N/S-RdRP* Ct values among non-VOCs, B.1.1.7, and B.1.351 FIGURE 1A, TABLE 1. In all clades except B.1.351, the *N* gene amplified earlier, with left-shifted Ct *N* vs Ct *S/RdRP* and Ct *E* ($P < .05$). B.1.351 was distinct from all other clades, showing a right-shifted (delayed) *N* gene amplification in the Allplex SARS-CoV-2 Assay. In the independent validation cohort (n = 153), which also contained 7 P.1 isolates, this delayed *N* gene amplification was confirmed as a characteristic feature of the B.1.351 lineage FIGURE 1B, TABLE 1. Correlation analyses of Ct values indicated an identical amplification profile of *E* and *S/RdRP* in all strains FIGURE 1C, TABLE 2 and graphically illustrate that the delayed *N* amplification vs both *S/RdRP* FIGURE 1D and *E* FIGURE 1E in the B.1.351 lineage is clearly proportional (slope = 1.21; 95% CI, 1.13-1.28) for *N* vs *E* TABLE 2, indicating a gradually more pronounced *N* gene amplification delay at lower viral loads.

Combined Use of Ct Values for Detection of the B.1.351 Lineage: VOC.V2 Probability Score

Next, logistic regression analysis was conducted to investigate whether Ct values could predict the presence of the B.1.351 lineage in individual patient samples in the study cohort, resulting in a fast and inexpensive screening tool. In a model that included all 3 PCR targets, Ct *E* ($P < .0001$) and Ct *N* ($P < .0001$) but not *S/RdRP* ($P = .5662$) independently contributed to prediction of B.1.351 infection. Repeating the regression analysis with only Ct *E* and Ct *N* indicated that a lower Ct

Downloaded from https://academic.oup.com/ajcp/article/157/5/731/6415253 by Hasselt University user on 01 June 2022

© American Society for Clinical Pathology

HTTPS://DOI.ORG/10.1093/AJCP/AQAB186

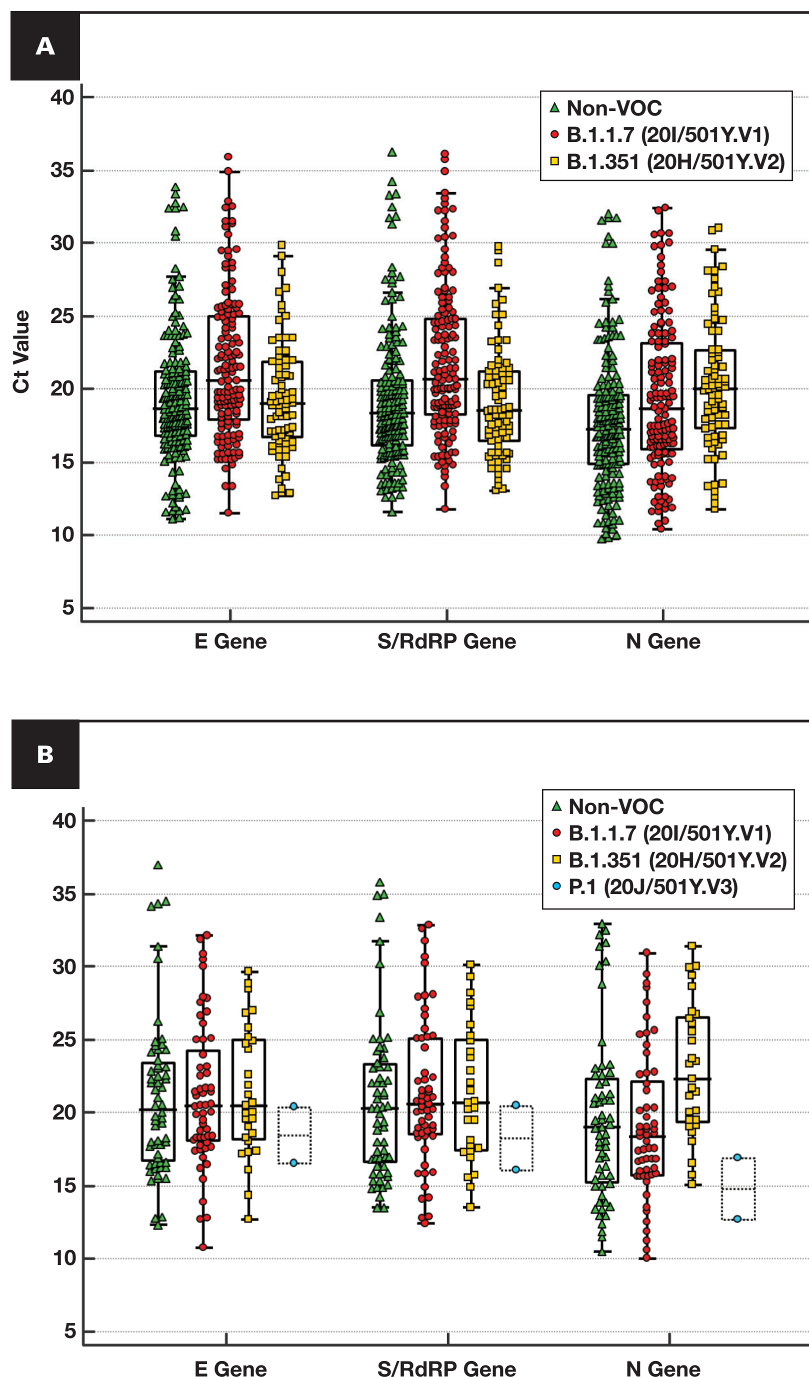

FIGURE 1 Distribution of Allplex SARS-CoV-2 Assay (Seegene) *E/N/S-RdRP* cycle threshold (Ct) values in B.1.1.7, B.1.351 and all non–variant of concern (VOC) lineages combined. **A, B**, Plotted distribution of raw Ct values (x-axis) of the *E, S/RdRP,* and *N* gene targets in the study (**A**, n = 397) and validation (**B**, n = 153) cohorts in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) lineages genotyped by whole-genome sequencing: B.1.1.7, B.1.351, P.1 (only present in the validation cohort) and all other, non-VOC lineages combined.

E value and a higher Ct *N* value had strong diagnostic power to detect B.1.351 (AUC = 0.94; 95% CI, 0.91-0.96). It allowed the calculation of a VOC.V2 probability score (p) for each individual sample using the following formula: $\ln(p/[1 - p]) = 0.9449 + 1.44091 \times CtNgene - 1.51860 \times Ct_{Egene}$ (Supplemental Table 3). The distribution of this VOC.V2 probability score for all non-VOC clades, B.1.1.7, and B.1.351 is plotted in FIGURE 2A. As expected, B.1.351 isolates had significantly higher VOC.V2 probability scores ($P < .0001$), although partially overlapping with other lineages.

The VOC.V2 probability score was further investigated as an independent single test result and its diagnostic power for individual patient samples evaluated. Remote operator characteristic analysis FIGURE 2B showed an AUC of 0.94 (95% CI, 0.91-0.96), with an optimal cut-point of 0.12 TABLE 3 for dichotomous test

Downloaded from https://academic.oup.com/ajcp/article/157/5/731/6415253 by Hasselt University user on 01 June 2022

HTTPS://DOI.ORG/10.1093/AJCP/AQAB186
© American Society for Clinical Pathology

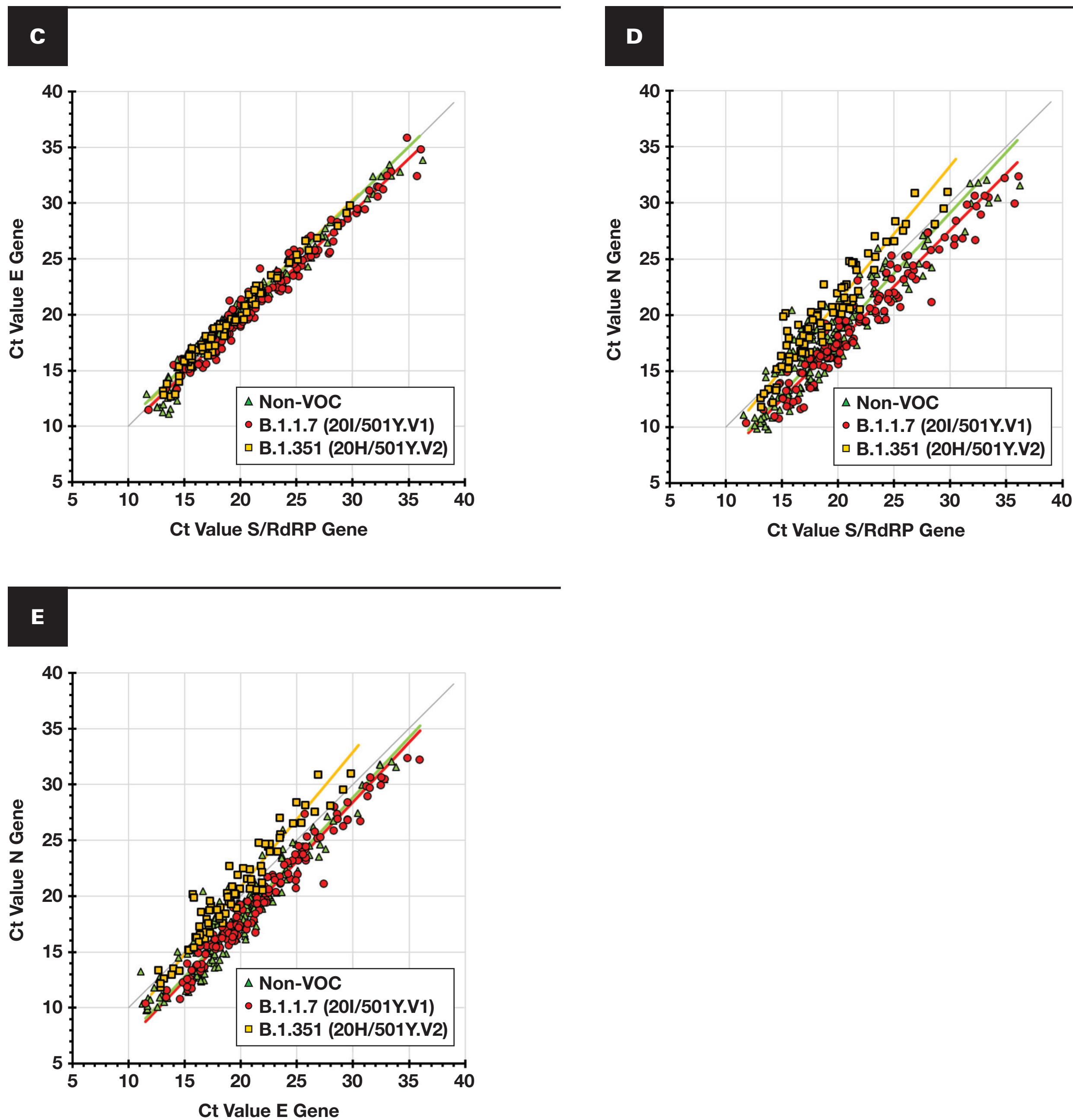

FIGURE 1 (cont) **C-E**, Correlation analysis (see also TABLE 2) of Ct values of the *E* vs *S/RdRP* channel (**C**), the *N* vs *S/RdRP* channel (**D**), and the *N* vs *E* channel (**E**) in the indicated SARS-CoV-2 lineages.

use, resulting in a sensitivity of 98.7% (95% CI, 93.1%-100.0%) and specificity of 79.9% (95% CI, 75.1%-84.2%) for screening of the B1.351 lineage. At a prevalence (pretest probability) of 19.6% of B.1.351 in our study cohort, this process resulted in a high negative predictive value (NPV) of 99.6% and a positive predictive value (PPV) of 54.6% TABLE 3.

Calculation of LRs for different VOC.V2 probability score result intervals showed that the probability of SARS-CoV-2 B.1.351 variant increased with increasing score FIGURE 2C, TABLE 3. A score of 0.12 or less almost completely ruled out B.1.351, with an LR of 0.02 (95% CI, 0.00-0.11), while scores above 0.50 strongly increased the probability, with LRs of 12.01 (95% CI, 7.21-20.02). Relations of post-test and pretest probabilities for the proposed VOC.V2 probability score intervals are shown in FIGURE 2D, illustrating the score's power to detect B.1.351 and taking the prevailing prevalence of this lineage in the population as a pretest probability. The clinical utility of the LR is summarized in a diagnostic algorithm in FIGURE 2E.

Validation of the VOC.V2 Probability Score in an Independent Cohort

The VOC.V2 probability score calculated from the study cohort was then tested in an independent validation cohort of 153

Downloaded from https://academic.oup.com/ajcp/article/157/5/731/6415253 by Hasselt University user on 01 June 2022

© American Society for Clinical Pathology

TABLE 1 Cycle Threshold Characteristics of Distinct SARS-CoV-2 Lineages on the Allplex SARS-CoV-2 Assay[a]

Patient Group	Non-VOC	B.1.1.7 (20I/501Y.V1)	B.1.351 (20H/501Y.V2)	P.1(20J/501Y.V3)
Study cohort (n = 397)				
No. (%)	178 (44.8)	141 (35.5)	78 (19.6)	–
Ct *E* gene, median (IQR)	18.7 (16.8-21.2)	20.6 (18.0-24.9)[b]	19.1 (16.8-21.9)[c]	–
Ct *S/RdRP* gene, median (IQR)	18.4 (16.2-20.6)	20.7 (18.3-24.8)[b]	18.5 (16.5-21.3)[c]	–
Ct *N* gene, median (IQR)[d]	17.3 (14.9-19.6) [d,e]	18.6 (15.9-23.1)[b,d,e]	20.1 (17.4-22.7)[f]	–
Validation cohort (n = 148)				
No. (%)	60 (39.2)	57 (37.3)	29 (19.0)	7 (4.6)
Ct *E* gene, median (IQR)	20.2 (16.8-23.4)	20.5 (18.1-24.2)	20.5 (18.2-25.0)	18.6 (18.2-23.3)
Ct *S/RdRP* gene, median (IQR)	20.3 (16.6-23.3)	20.6 (18.5-25.0)	20.6 (17.5-25.0)	18.4 (17.6-22.2)
Ct *N* gene, median (IQR)	19.0 (15.3-22.4)	18.4 (15.8-22.1)[d,e]	22.3 (19.4-26.5)[c,f,g]	17.0 (15.8-21.9)

Ct, cycle threshold; IQR, interquartile range; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; VOC, variant of concern.
[a]The table lists the prevalences of the indicated SARS-CoV-2 clades in the study cohort and validation cohort and their median (IQR, 25th-75th percentiles) Ct values on the Allplex SARS-CoV-2 Assay (Seegene). From left to right: all non-VOC lineages combined (for the detailed composition, see Supplementary Table 1), the B.1.1.7 UK variant (20I/501Y.V1 clade, VOC.V1), the South African B.1.351 lineage (20H/501Y.V2 clade, VOC.V2), and the Brazilian P.1 lineage (20J/501Y.V3, VOC.V3). Letters indicate corresponding statistical differences as calculated by Mann-Whitney *U* testing.
[b]Indicates differences of VOC.V1 with non-VOCs for which $P < .05$ was considered statistically significant.
[c]Indicates differences of VOC.V2 with VOC.V1 for which $P < .05$ was considered statistically significant.
[d]Indicates differences of the Ct *N* gene with the Ct *E* gene for which $P < .05$ was considered statistically significant.
[e]Indicates differences of the Ct *N* gene with the Ct *S/RdRP* gene for which $P < .05$ was considered statistically significant.
[f]Indicates differences of VOC.V2 with non-VOCs for which $P < .05$ was considered statistically significant.
[g]Indicates differences of VOC.V2 with VOC.V3 for which $P < .05$ was considered statistically significant.

consecutive samples analyzed by WGS and PCR. This validation cohort contained a prevalence TABLE 1 similar to B.1.1.7 (39.2%) and B.1.351 (19.0%) but also contained 7 (4.6%) P.1 isolates. Overall, diagnostic performance was identical, with a 100% NPV and 53.7% PPV using the optimal cut-point of 0.12 and increasing probability of B.1.351 screening with increasing VOC.V2 probability score, attaining an LR of 17.82 (95% CI, 8.05-39.41) at scores above 0.50.

Integration of the VOC.V2 Probability Score With Reflex PCR for N501Y and 69.70del for Rapid, Cost-effective VOC Detection

A subset (n = 172) of study and validation cohort samples was additionally analyzed by targeted PCR for N501Y and 69.70del mutations in the spike protein (Supplemental Table 2). As shown in FIGURE 3A, the combined use of the *N* amplification delay in the basic *E/N/S-RdRP* PCR assay with these 2 biologically relevant mutations allowed for a rapid and cost-effective discrimination of the 3 VOCs in our population (B.1.1.7, B.1.351, and P.1). All 3 share the N501Y mutation. The additional presence of 69.70del indicates B.1.1.7. Strains carrying N501Y but not 69.70del can be discerned as either B.1.351 or P.1 based on VOC.V2 probability score above or below 0.12, respectively. FIGURE 3B shows the concordance of this simplified diagnostic approach vs the gold standard of WGS, with excellent agreement (weighted $\kappa = 0.987$ [95% CI, 0.969-1.000]).

Multicenter Validation

To evaluate the interlaboratory robustness of the VOC.V2 probability score, an additional validation was performed using the Allplex SARS-CoV-2 Assay Ct *E/S-RdRP/N* data obtained from an independent laboratory that used an alternative RNA-extraction setup. In a cohort containing 308 consecutive SARS-CoV-2 PCR-positive samples (n = 68 B.1.351, n = 73 non-VOC lineages, and n = 167 B.1.1.7 isolates as measured by WGS) (Supplemental Table 4), the VOC.V2 probability score achieved a sensitivity of 100% (95% CI, 94.7%-100%) at 60.8% (95% CI, 54.4%-67.1%) specificity using binary thresholding with a single cutoff of 0.12. All B.1.351 isolates showed a VOC.V2 probability score above 0.5, a cutoff associated with an LR of 12.6 (95% CI, 8.2-19.5).

Time Sensitivity of Ct Modeling

The VOC.V2 probability score was optimized to screen for B.1.351 in an epidemiological context with the prevalence of several viral lineages, as described in TABLE 1. The subsequent rise to dominance of B.1.1.7 and, more recently B.1.617.2 (the Delta variant), with outcompetition of B.1.351, evidently affects this time sensitive modeling. Virtual implementation of the VOC.V2 probability score in a context dominated by B.1.617.2 indicates that VOC.V2 probability scores of B.1.617.2 isolates (median, 0.37 [interquartile range (IQR), 0.12-0.55]; n = 482) are significantly ($P < .0001$) lower than those of B.1.351 (median, 0.77 [IQR, 0.60-0.91] in the validation cohort [n = 29]). However, the extensive overlap of their VOC.V2 probability score distribution clearly precludes their confident distinction at an individual sample level in the hypothetical scenario of their co-occurrence (Supplemental Table 5, Supplemental Figure 1) and underlines the time sensitivity of the actual modeling and the conceptual need for continuous adaptation as a function of shifts in lineage prevalences.

DISCUSSION

This article describes a VOC.V2 probability score calculated by regression modeling of the Ct values of the widely used Allplex

HTTPS://DOI.ORG/10.1093/AJCP/AQAB186
© American Society for Clinical Pathology

Downloaded from https://academic.oup.com/ajcp/article/157/5/731/6415253 by Hasselt University user on 01 June 2022

TABLE 2 Correlation and Regression Analysis of Cycle Threshold Values for the *E* Gene, *S/RdRP* Gene, and *N* Gene Within Different SARS-CoV-2 Lineages in the Study Cohort[a]

Variant Strain	Ct *E* Gene (Y) vs Ct *S/RdRP* Gene (X)			Ct *N* Gene (Y) vs Ct *S/RdRP* Gene (X)			Ct *N* Gene (Y) vs Ct *E* Gene (X)		
	r_s (95% CI)	Slope (95% CI)	Intercept (95% CI)	r_s (95% CI)	Slope (95% CI)	Intercept (95% CI)	r_s (95% CI)	Slope (95% CI)	Intercept (95% CI)
Non-VOC (n = 178)	0.99 (0.98-0.99)	0.98 (−1.08 to 1.08)	0.78[b] (0.21-1.30)	0.88 (0.85-0.91)	1.08 (1.00-1.15)	−3.17[b] (−4.54 to −1.71)	0.91 (0.88-0.93)	1.07[c] (1.02-1.13)	−3.27[b] (−4.41 to −2.19)
B.1.1.7 (20I/501Y.V1) (n = 141)	0.98 (0.97-0.99)	0.96[c] (0.93-0.99)	0.33 (−0.26 to 0.96)	0.97 (0.96-0.98)	1.00 (0.96-1.05)	−2.56[b] (−3.66 to −1.63)	0.98 (0.98-0.99)	1.06[c] (1.03-1.10)	−3.46[b] (−4.29 to −2.74)
B.1.351 (20H/501Y.V2) (n = 78)	0.99 (0.98-0.99)	0.99 (0.95-1.03)	0.52 (−0.35 to 1.28)	0.93 (0.90-0.96)	1.21[c] (1.11-1.32)	−3.05[b] (−4.78 to −1.11)	0.95 (0.92-0.97)	1.21[c] (1.13-1.28)	−3.26[b] (−4.53 to −1.77)

CI, confidence interval; Ct, cycle threshold; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; VOC, variant of concern.
[a]Correlation analysis of Ct *E* gene (left panel) and *N* gene (right panel) vs the combined Ct of *S/RdRP* targets in the study cohort (n = 397) indicates that the *N* gene amplification delay in the B.1.351 lineage is proportional, with more pronounced *N* gene amplification delay at a higher Ct (lower viral load) (slope = 1.21 with confidence interval, excluding 1).
[b]Significant systematic difference.
[c]Significant proportional difference.

SARS-CoV-2 Assay. As a stand-alone test, this derived score can screen for infection by the South African B.1.351 (20H/501.V2) VOC with a sensitivity approaching 100% at an acceptable specificity of approximately 80%. Its high sensitivity thus ensures that all relevant samples are selected for subsequent reflex testing by mutation-specific PCRs or WGS, 2 tests that achieve near-maximal specificity.

The methodological strengths of our study are (1) its use of independent training and validation cohorts, both sufficiently sized and consisting of consecutive and representative samples; (2) the use of WGS as the gold standard to unambiguously determine clinical truth (exact viral lineage); and (3) the subsequent validation of our VOC.V2 probability score on PCR data obtained from an independent laboratory with a different RNA-extraction setup, highlighting that the model is robust and can also be implemented on Allplex SARS-CoV-2 Assay Ct data from other labs generated by other types of liquid handlers or thermal cyclers.

Limitations of our study are, first, that its application for B.1.351 screening is transferable only to labs that use the Allplex SARS-CoV-2 Assay, similar to the limitation of the SGFT feature in the TaqPath assay. Second, this application is time sensitive: the relatively high mutation rate of coronaviruses results in a rapid evolutionary rate. During preparation of the present report, both the B.1.1.7 and B.1.351 lineages have subsided and have been replaced by the B.1.617.2 lineage as the dominant strain, rendering B.1.351 screening obsolete.

The possible impact of our findings on SARS-CoV-2 surveillance is predominantly conceptual and situated at epidemiological level 1. The VOC.V2 probability score provided an immediate suggestion for an emerging VOC (B.1.351) based on the Ct results of the standard PCR (typically within 12 to 24 hours after sample reception), thus avoiding a delay of 24 hours to several days compared with mutation-specific PCR and WGS, respectively. This timing had no direct impact on individual patient management. It is also uncertain if this rapid screening assisted in the regional containment of B.1.351 because this strain was clearly outcompeted by increasingly infectious variants—first B.1.1.7, and then B.1.617.2. Yet, scores such as our VOC.V2 probability score can be easily implemented on any multitarget SARS-CoV-2 PCR assay and, once modeled on any given assay or platform, can be exchanged with other users of this same assay or platform at no additional cost. The underlying *N* gene amplification delay in B.1.351 isolates in the widely used Allplex SARS-CoV-2 Assay is thus comparable to a similarly informative artefact, the SGFT in the equally widely used TaqPath PCR assay. The latter formed the basis of a cost-effective 2-step screening approach for B.1.1.7 implemented in many countries and yielding dense epidemiological data.[17,18]

Conceptually, our study shows that additional relevant diagnostic information can be extracted from raw PCR data and outlined an approach that can be iterated for other PCR assays, provided that a lab has access to a gold standard method for viral genotyping. Our study thus provides a straightforward framework for similar modeling on other PCR platforms for early detection of newly emerging viral strains. It provides a fast and cost-effective screening tool for the concerning B.1.351 SARS-CoV-2 variant based on secondary use of data from a

Downloaded from https://academic.oup.com/ajcp/article/157/5/731/6415253 by Hasselt University user on 01 June 2022

© American Society for Clinical Pathology

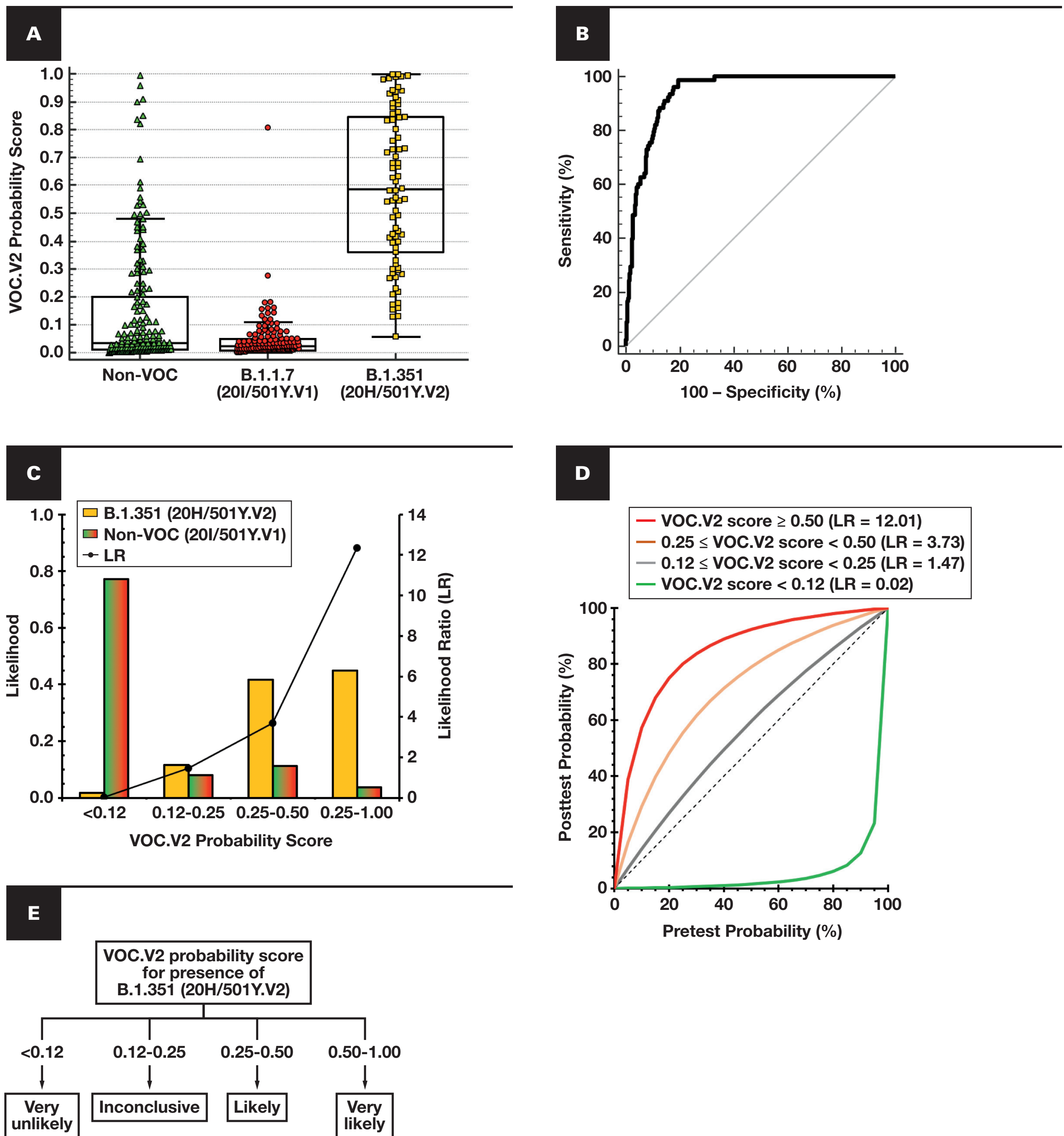

FIGURE 2 Diagnostic performance of the variant of concern (VOC).V2 probability score for detection of the B.1.351 variant vs all other lineages. **A**, Distribution of calculated VOC.V2 probability scores of non-VOC samples combined (n = 178), B.1.1.7 samples (n = 141), and B.1.351 samples (n = 78) in the study cohort. **B**, Receiver operating characteristic curve of the VOC.V2 probability score for detection of B.1.351 vs all other lineages combined (see also TABLE 3). **C**, Likelihood (left y-axis) of the indicated result intervals of the VOC.V2 probability scores given the status of B.1.351 vs the status of non-VOCs or B.1.1.7 combined. Likelihood ratios (LRs) are shown on the right y-axis. **D**, Posttest probability for B.1.351 as a function of pretest probability and of the result interval of the VOC.V2 probability score. **E**, Summary of the diagnostic algorithm of the VOC.V2 probability score for screening of B.1.351.

Downloaded from https://academic.oup.com/ajcp/article/157/5/731/6415253 by Hasselt University user on 01 June 2022

© American Society for Clinical Pathology

TABLE 3 Diagnostic Performance of the VOC.V2 Probability Score in the Study and Validation Cohorts Using Single Cutoff or Multiple Result Intervals[a]

VOC.V2 Probability Score	B.1.351	All Other Lineages Combined	LR (95% CI)	Sn, % (95% CI)	Sp, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Accuracy, % (95% CI)
Study cohort (n = 397)	n = 78	n = 319	–	–	–	–	–	–
Single cutoff (dichotomous)								
<0.12	1	255	0.02 (0.00-0.11)	98.7 (93.1-100.0)	79.9 (75.1-84.2)	54.6 (46.0-63.0)	99.6 (97.8-100.0)	83.6 (79.7-86.9)
≥0.12	77	64	4.92 (3.95-6.13)					
Multiple result intervals								
0.00-0.12	1	255	0.02 (0.00-0.11)	–	–	–	–	–
0.12-0.25	9	25	1.47 (0.72-3.03)					
0.25-0.50	21	23	3.73 (2.18-6.39)					
0.50-1.00	47	16	12.01 (7.21-20.02)					
Validation cohort (n = 153)	n = 29	n = 124	–	–	–	–	–	–
Single cutoff (dichotomous)								
<0.12	0	99	0.00 (0.00-0.34)	100.0 (88.1-100.0)	79.8 (71.7-86.5)	53.7 (39.6-67.4)	100.0 (96.3-100.0)	83.6 (77.0-88.7)
≥0.12	29	25	4.96 (3.49-7.04)					
Multiple result intervals								
0.00-0.12	0	99	0.00 (0.00-0.34)	–	–	–	–	–
0.12-0.25	1	11	0.39 (0.05-2.89)					
0.25-0.50	3	8	1.60 (0.45-5.68)					
0.50-1.00	25	6	17.82 (8.05-39.41)					

CI, confidence interval; LR, likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; Sn, sensitivity; Sp, specificity; VOC, variant of concern.
[a]Diagnostic performance of the VOC.V2 probability score as a separate test result for detection of the B.1.351 lineage compared with all other lineages combined (non-VOC, B.1.1.7, and P.1) was assessed using a dichotomous single-cutoff approach or by using LRs (LR, 95% CI) for the indicated multiple result intervals in the study (upper panel) and validation (lower panel) cohorts. For dichotomous use, from left to right: Sn, Sp, PPV, NPV with 95% CI. PPV and NPV based on the observed prevalences of B.1.351 in study and validation cohorts respectively.

Downloaded from https://academic.oup.com/ajcp/article/157/5/731/6415253 by Hasselt University user on 01 June 2022

© American Society for Clinical Pathology

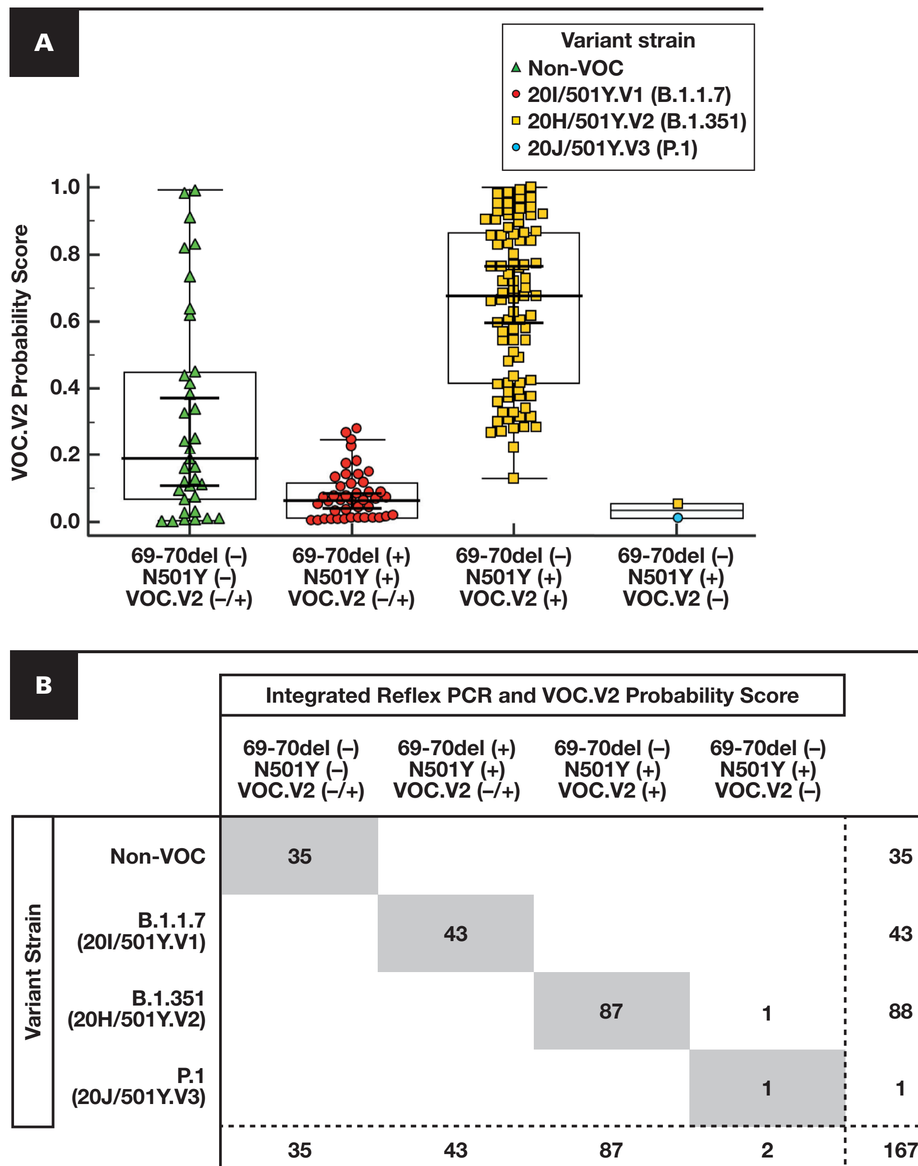

FIGURE 3 Integration of the variant of concern (VOC).V2 probability score with N501Y and 69.70 deletion (del) mutation–specific reflex polymerase chain reaction (PCR) for differentiation of severe acute respiratory syndrome coronavirus 2 VOCs 1 to 3. **A**, Distribution of the VOC.V2 probability score (y-axis) in B.1.1.7; B.1.351; P.1 (only present in the validation cohort); and all other, non- VOC lineages combined in a subset of samples for which mutation-specific reflex PCR was performed for the spike *N501Y* and 69.70del mutations. **B**, Concordance analysis of an integrated PCR analysis of the VOC.V2 probability score and N501Y and 69.70del mutations (horizontal) compared with whole-genome sequencing genotyping (vertical) as the gold standard in a subset of 172 samples. Weighted κ = 0.993 (95% confidence interval, 0.979-1.000).

widely used SARS-CoV-2 PCR. More generally, it outlines a straightforward strategy to create greater diagnostic value from available data.

REFERENCES

1. Davies NG, Abbott S, Barnard RC, et al; CMMID COVID-19 Working Group; COVID-19 Genomics UK (COG-UK) Consortium. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science.* 2021;372:eabg3055.
2. Volz E, Mishra S, Chand M, et al; COVID-19 Genomics UK (COG-UK) Consortium. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature.* 2021;593:266-269.
3. Collier DA, De Marco A, Ferreira IATM, et al; CITIID-NIHR BioResource COVID-19 Collaboration; COVID-19 Genomics UK (COG-UK) Consortium. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature.* 2021;593:136-141.
4. Supasa P, Zhou D, Dejnirattisai W, et al. Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell.* 2021;184:2201-2211.e7.

© American Society for Clinical Pathology

Downloaded from https://academic.oup.com/ajcp/article/157/5/731/6415253 by Hasselt University user on 01 June 2022

5. Hodcroft EB. CoVariants. CoVariants Web site. https://covariants.org/. Accessed March 30, 2021.
6. Davies NG, Jarvis CI, Edmunds WJ, et al; CMMID COVID-19 Working Group. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature.* 2021;593:270-274.
7. Challen R, Brooks-Pollock E, Read JM, et al. Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ.* 2021;372:n579.
8. Wang Z, Schmidt F, Weisblum Y, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature.* 2021;592:616-622.
9. Liu Y, Liu J, Xia H, et al. Neutralizing activity of BNT162b2-elicited serum. *N Engl J Med.* 2021;384:1466-1468.
10. Muik A, Wallisch AK, Sänger B, et al. Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by BNT162b2 vaccine-elicited human sera. *Science.* 2021;371:1152-1153.
11. Tegally H, Wilkinson E, Giovanetti M, et al. Emergence of a SARS-CoV-2 variant of concern with mutations in spike glycoprotein [published online ahead of print March 9, 2021]. *Nature.* doi:10.1038/s41586-021-03402-9.
12. Faria NR, Mellan TA, Whittaker C, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science.* 2021;372:815-821.
13. Sabino EC, Buss LF, Carvalho MPS, et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet.* 2021;397:452-455.
14. Wu K, Werner AP, Moliva JI, et al. mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants [preprint published online ahead of print January 25, 2021]. *bioRxiv.* doi:10.1101/2021.01.25.427948.
15. Wibmer CK, Ayres F, Hermanus T, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med.* 2021;27:622-625.
16. Madhi SA, Baillie V, Cutland CL, et al; NGS-SA Group; Wits-VIDA COVID Group. Efficacy of the ChAdOx1 nCoV-19 COVID-19 vaccine against the B.1.351 variant. *N Engl J Med.* 2021;384:1885-1898.
17. Bal A, Destras G, Gaymard A, et al; COVID-Diagnosis HCL Study Group. Two-step strategy for the identification of SARS-CoV-2 variant of concern 202012/01 and other variants with spike deletion H69-V70, France, August to December 2020. *Euro Surveill.* 26;2021;26:2100008.
18. Chand M, Hopkins S, Dabrera G, et al. Investigation of novel SARS-CoV-2 variant: variant of concern 202012/01. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959438/Technical_Briefing_VOC_SH_NJL2_SH2.pdf. Accessed March 30, 2021.

Downloaded from https://academic.oup.com/ajcp/article/157/5/731/6415253 by Hasselt University user on 01 June 2022

© American Society for Clinical Pathology