Best Practice Guidance: Recommendations for *in silico* analysis of SARS-CoV-2 sequences to identify potential negative impact of genetic variants on molecular diagnostic assay performance

Authors & affiliations: Nathaniel Storey[1], Jim Huggett[2], Rui Pereira[2], Denise O'Sullivan[2], Kathryn Harris[1], Rebecca Riches-Duit [3], Louise Oliver[4], Richard Vipond[4], Kirsten Kelly[4], Monee Shamsher[4], Lisa Stevens[4]

1. Department of Microbiology, Virology and Infection Prevention & Control, Great Ormond Street Hospital NHS Foundation Trust.
2. Molecular and Cell Biology Team, National Measurement Laboratory (LGC)
3. Medicines and Healthcare products Regulatory Agency (MHRA)
4. UK Health and Security Agency (UKHSA)

Issue date: 04 August 2023

Contents:

1. Scope

This document outlines the procedure for manufacturers of molecular tests, limited to nucleic acid amplification tests (NAATs) such as polymerase chain reaction (PCR) assays, as part of post-market surveillance requirements to routinely monitor for new genetic mutations and variants currently circulating within the United Kingdom, and to assess the impact of the associated genetic changes, mutations, or variants on the performance of their assay.

2. Introduction

SARS-CoV-2 genetic variants exist that have the potential to lead to phenotypic changes manifesting as differences in viral infectivity, burden, and tropism during the course of infection. These could be advantageous or deleterious to a diagnostic pipeline. Consequently, any diagnostic solution (antigen, serology, biomarker, or molecular diagnostic) should ideally empirically monitor the impact of significant variants during routine test evaluation. As molecular diagnostic assays directly target the genetic sequence, any changes to those sequences in a variant have the potential to disrupt their performance.

NAATs are tests that use primers and probes (short oligonucleotides that complement the target sequence of interest) to bind and detect specific virus RNA target sequences. While some NAATs detect RNA directly, methods like PCR, loop-mediated isothermal amplification (LAMP) and recombinase polymerase amplification (RPA) detect DNA and require reverse transcription (RT) to perform. For a SARS-CoV-2 diagnostic assay they are designed to complement the SARS-CoV-2 genome. Several Variants of Concern (VOC) and Variants under investigation (VUI) have been recognised in the SARS CoV-2 virus that are circulating globally. Viral variants can cause genetic changes which can impact primer/probe sequences. This occurs because the variant sequences will no longer perfectly complement the primer/probe sequence intended to be detected. This can have no effect, a marginal effect, or a catastrophic effect on the performance potentially producing a false negative result.

This problem can be mitigated by *in silico* sequence assessment. However, two pieces of information are required to enable this:

1. The genomic sequence of the genetic variants in question.
2. The sequence of the genetic region targeted by the molecular assay being used (typically small DNA sequences (oligonucleotides) known as primers or probes).

Armed with the information above it is relatively easy to compare a molecular assay with circulating variants and determine whether any genetic change within the primes or probes may impact on a molecular diagnostic test. Furthermore, using suitable criteria, an assessment can be made as to whether the change may be minor or result in catastrophic failure of the diagnostic test leading to false negative results or gene dropouts and feed into an early warning and response system. This assessment also provides crucial information for the design of the test; if the genetic change impacts test performance, then the test manufacturers may need to redesign the primer or probe sequences and bring the test to required sensitivity/specificity.

Under regulation 34A of the Medical Devices Regulations 2002 no antigen or molecular detection COVID-19 (SARS-CoV-2) test may be placed on the UK market without first being validated against minimum performance standards through a Coronavirus Test Device Approvals (CTDA) desktop review (COVID-19 test approval: how to apply - GOV.UK (www.gov.uk)). If the primers or probes have to be re-designed to consider a potential issue with a VOC or VUI, this assay may need to be re-validated and submitted to the CTDA for approval before it can be sold on the UK market. Tests listed on the CTDA register are approved for the Instructions for Use (IFU) version listed on the register only. Where a manufacturer updates their IFU they shall submit the amended IFU to CTDA for assessment as a minor or major change. If deemed a major change a new full application will need to be submitted. As CTDA is a goods regulation, only those tests (IFU versions) listed on the register are allowed to be sold on the UK market.

If a test is re-designed, changed, or an IFU is updated, to retain UKCA/CE marking the manufacturer may also need to perform a conformity assessment and possibly gain approval from its Approved Body or Notified Body, depending upon the classification of the IVD under the UK MDR 2002. To address this change to the IVD or to correct another problem, the manufacturer may also need to submit a Field Safety Notice (FSN) and Field Safety Corrective Action (FSCA) form to the MHRA, and to send the FSN to their customers to alert them of any changes and of corrective actions to mitigate risk. Please refer to section 8 for further detail on FSNs.

In this document we recommend a methodology for *in silico* analysis of sequence variation in the genetic regions targeted by reverse transcriptase quantitative polymerase chain reaction (RT qPCR).

Separate guidance has been produced tailored to antigen tests, which can be found on the MHRA webpage: Guidance for manufacturers: diagnostic assurance with SARS-CoV-2 variants in circulation.

3. Frequency recommendations for *in silico* analysis

The frequency of running *in silico* analysis is of immense importance. Due to the rapidly evolving distribution of SARS-CoV-2 lineages both within the UK and globally, *in silico* analysis should be conducted afresh every two months and in a timely manner. The primer or probe sequences should be compared against a time span of 60 days of variant sequence submissions circulating in the UK within the chosen database (see section "4. Database and variant considerations"), unless a new emerging variant requires immediate investigation – in this instance you will be notified by the MHRA. This would allow rapid identification of lineages that may be affected by the genetic changes within assay target regions and allow for continuous monitoring for the stability of these genetic changes.

The prevalence over time can be skewed due to the lag time between sample collection, sequencing, and the submission of sequence data to a database. To address this uncertainty, a rolling window of 60 days is recommended. The frequency of analysis should remain the same to ensure comparability.

However, during periods of rapid expansion of specific lineages or increased diversity, more frequent and targeted analysis may be required.

4. Database and variant considerations

Sequencing data can be used to understand whether current diagnostic tests remain fit for purpose, as well as inform the development of new diagnostics. Considerations for the nucleotide database against which the primer or probe (target query) sequences are to be compared against should include the assessment of the quality of sequence deposited, diversity or number of sequences deposited for each lineage or variant, and whether the data is representative of UK lineages.

Sequence quality is of particular importance and care should be taken to reduce the number of aligned sequences containing unknown or ambiguous bases. The diversity or number of sequences deposited for each variant will vary depending on the database used so it is recommended that a global database such as GISAID (GISAID - gisaid.org) or country specific, in this case representing UK specific lineage data from COVID-19 Genomics UK Consortium (COVID-19 Genomics UK Consortium (cogconsortium.uk), be used. These sequences can be downloaded from databases in any format that will allow for the analysis and sorting of the data. UK specific sequence data is available either as a single file from the COG-UK data page or international data can be searched with specific criteria from GISAID. Under normal circumstances it may be beneficial to filter sequences by using the 'complete' definition of GISAID. However, in certain circumstances such as a rapidly emerging variant with significant divergence from previous lineages, we recommend that none of the filters are selected in order to obtain a suitable quantity of sequences for analysis. Regular updates on SARS-CoV-2 lineage prevalence and growth rates observed in the UK can be obtained from the GOV UK website (SARS-CoV-2: genome sequence prevalence and growth rate - GOV.UK (www.gov.uk)) and manufacturers should use the latest update to determine which VOCs/VUIs sequences to interrogate. Up to date information about SARS-CoV-2 genome mutations and variants of interest circulating in the UK can also be accessed through the COG-UK Mutation Explorer (COG-UK-ME)

interface COG-UK/Mutation Explorer (gla.ac.uk). This latter source is updated twice a week (correct as of July 2023).

Although these may not have been heightened to VOC/VUI status, it is prudent for manufacturers to track the variants here and other variant tracking sources such as Tracking SARS-CoV-2 variants (who.int) as variants arising globally may have an impact on circulating UK variants through international travel.

The target query sequences should be compared to all suitable sequences for the VOC or VUI within a time span of 60 days of variant sequence submissions to the database. In instances where the number of sequences available for a representative lineage circulating within the UK is limited, sequences should be analysed to yield significant and reliable interpretation of the data. If the sequence data is limited by low sequence numbers (<100 sequences), this should be caveated in the interpretation of the results. In the latter instance where the analysis has been conducted with limited number of sequences, the manufacturer should explain this in a caveat to highlight the higher frequency of mutations (if any) for these lineages and should continue to monitor this lineage in the next analysis.

5. Bioinformatics recommendations

The analysis of sequences should be performed by a trained bioinformatician or person with significant experience in the handling of large databases and the tools used should be both peer-reviewed and version controlled. The version information and references of all tools and software should be recorded for each analysis and the use of differing tools or versions between analyses is strongly discouraged.

For use as the target query sequence, it is recommended that short sequences (<200bp) are extracted from complete records that correspond to regions associated with the target nucleotide sequence. This reduces the chances of non-specific alignment and allows the region of interest to be more accurately quality controlled. Any sequence containing unknown or ambiguous bases within the target query region should be removed from the database to prevent misidentification of mutations. In the GISAID database, the hCoV-19/Wuhan/WIV04/2019 (WIV04) GISAID - hCoV-19 Reference Sequences is the official reference sequence employed (EPI_ISL_402124). WIV04 was chosen because of its high-quality genome sequence and because it represented the consensus of a handful of early submissions for the beta coronavirus responsible for COVID-19. On the protein level NCBI Reference Sequence: NC_045512.2 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan is the predominantly used reference sequence.

From full length variant sequences extracted from the appropriate database(s), multiple sequence alignments (MSA) or pairwise alignments should be performed to identify nucleotide changes within target sequences, using peer-reviewed tools. For generating MSA from large sequence databases manufacturers should use an aligner such as minimap2 (https://github.com/lh3/minimap2), in order to reduce the computational requirements, it is recommended that multiple pairwise alignments are performed on consensus sequences rather than MSA-proper. From the sam files produced by the aligner of choice, further tools such as gofasta (https://github.com/virus-evolution/gofasta) may be used to convert the data into fasta format for further analysis. For the specific analysis of amino acid changes within SARS-CoV-2 genes, the use of tools such as Nextalign (https://hub.docker.com/r/nextstrain/nextalign) may offer improvements in ease of use and is available in a containerised form.

Any lineages with previously identified mutations shall be continually updated within the database so that any further acquired mutations that may result in assay failure are rapidly identified.

Information of either or all sequence identifiers (for example, COG-UK ID or Accession No.) of sequences used in the analysis, or sufficient metadata (for example, lineage and epi week/date range) and sequence selection criteria to ensure that the data used can be independently identified, should be made available to the MHRA if requested.

Whilst some details on specific reagents, primer/probe, or other oligonucleotide sequences and how these are prepared are intellectual property and/or commercially sensitive information, this can be shared with the relevant competent authority. In the UK, this is the MHRA; such confidential information shall be part of the full data set required. The information is required so that the MHRA can monitor any issues that might occur with a given molecular test if a predominant variant arises.

6. Assessment of impact on molecular assay

If any mutations are identified in the target regions of primers and probes, several factors should be taken into consideration:

(1) the location of the sequence mismatch within the primers/probes:

Where a single genetic change occurs in the 5' (left side) of a primer, then the impact of a change will be reduced. However, as changes occur towards the 3' (right side) of the primer, then impact becomes more significant, potentially resulting in reduced analytical sensitivity and ultimately assay failure (a false negative test result). However, where differences occur at the location bound by the probe (also a short DNA molecule but used to detect the successful PCR reaction) this can lead to reduced analytical performance or failure, even if the primers are unaffected.

(2) multiple sequence mismatches within the primers/probes:

Where larger genetic changes occur, such as deletions or insertions, these will more likely result in catastrophic failure. Aggregates of mutations could impact a particular test to reduce performance from that previously established.

As a result of the *in silico* analysis performed, the frequency of mutations for each variant should be reported, and proportions calculated for regular surveillance purposes. A defined criteria to determine if the proportion of mutation(s) occurring can have an impact on the assay should be used. This resulting information from the *in silico* analysis should be formatted in an easy to review table and include the information below as a minimum:

1. Reference sequence used e.g., hCoV-19/Wuhan/WIV04/2019 (WIV04)
2. Location of target query sequence of primer/probe set analysed within the reference sequence.
3. Database used for mining of circulating sequences e.g., GISAID.
4. Timeframe when sequences were mined e.g., 01 March 2023 to 31 March 2023.
5. Database used to determine which variant sequences were analysed e.g., UK circulating variants.
6. Variant sequences analysed, using SARS-CoV-2 pangolin lineages as standard nomenclature e.g., BA.1
7. Location of mutation(s) within primer/probes on each variant e.g., A21830T.
8. Frequency and proportion of each mutation within each variant analysed.

The frequency and location of all mutations or mismatches within all assay targets shall be noted and percentage frequency based on the number of target sequences analysed should be calculated for regular surveillance purposes (for example, a single SNP/SNV should be recorded as: C21026T, 4.5% frequency).

9. Predicted impact of mutation(s) and what criteria is used for defining impact on the assay performance including a justification of the likelihood that the mutation may impact assay performance. A further justification for any actions taken, or not taken, based on the outcome of this analysis should also be mentioned.

Evidence of impact should be based on other predictions such as melting temperature (Tm) of the mutated nucleotide changes and assessed against predicted impact on the Ct values of the assay.

It is required that both favourable and unfavourable data should be reported to the MHRA as assurance of either positive or negative assay performance.

7. Wet laboratory tests

Details of any performance evaluation of tests against variants that have been conducted or how manufacturers plan to test to demonstrate performance with either recombinant *in vitro* transcribed RNA, g-blocks, or recombinant protein (containing the sequence changes), or characterised viral isolates, should be clearly communicated.

Where sequence variation is detected in primer or probe target sequences and this is predicted to impact assay performance, manufacturers shall conduct confirmatory wet lab testing of their product or assay against the variant sequence in question. It is advisable even if prior *in silico* analysis of variants does not identify any lineages carrying mutations in the target primer/probe sequences, these *in silico* predictions should still be challenged with a variety of wet lab evaluation.

Wet lab testing with a clinical sample with the mutation(s) within the primers/probe sequence, if available, is ultimately required to show its effect on assay performance. Where clinical material is not available, due to potential difficulties in identifying and acquiring such a sample, use of synthetic RNA targets with and without the mutation is acceptable. Limit of detection (LOD) testing should be performed in parallel for synthetic targets containing both the “wild-type” (no mutation) and mutation(s) representing the variant of interest.

Information from wet testing should be formatted in an easy to review table and include the information below as a minimum:

1. Type of material used in wet laboratory testing e.g., synthetic molecules (or recombinant protein) with defining mutations or clinical samples.
2. Methodology used to determine LOD.
3. LOD of the wild type (original reference sequence) as validated in IFU and when tested alongside variant sequences.
4. Number of samples used for wild type and variant sequences e.g., number of clinical samples used to conduct the experiment.
5. LOD of the variants using same standard units as used for wild type material.
6. Impact on the assay through wet testing.

If wet testing results show that the mutation negatively affects assay performance and risks the assay producing false negative results, then this shall be reported to the MHRA within 48 hours as a serious public health threat, and an FSN shall be issued to alert customers.

8. Field safety notice (FSN)

The UK medical devices regulations (UK MDR 2002) state that manufacturers must report incidents involving their device as soon as possible, and no later than 30 days after becoming aware of the incident. In line with UK MDR 2002 vigilance and field safety corrective action reporting requirements, the MHRA consider reports relating to SARS-CoV-2 Variants of Concern to be serious public health threats, therefore significant safety issues, including indirect patient harm through false negative and false positive results, should be reported within 48 hours. If the wet testing performance of any assay targets, including each oligonucleotide target sequence, is directly impacted by these new virus variant(s), a FSN shall be issued immediately to alert customers, and both the FSN and FSCA submitted to the MHRA. Similarly, if a manufacturer receives customer reports regarding false negative results, this shall be investigated immediately and reported to the MHRA. FSNs may also need to be issued if there are changes to the device IFU, or to the assay design, as field safety corrective actions may need to be undertaken. Further guidance on effective FSNs can be found at:

https://www.gov.uk/guidance/effective-field-safety-notices-fsns-guidance-for-manufacturers-of-medical-devices.

9. Post-market surveillance plan (PMSP)

Effective post-market surveillance shall be performed for SARS-CoV-2 screening and diagnostic assays under the UK MDR 2002 to continuously monitor, investigate, and assess the performance of an assay against newly emerging variants. A post-market surveillance plan shall be in place and this plan should be forwarded to the MHRA. It is also good practice to apply international standards to quality management systems (ISO 13485) and risk management (ISO 14971) for medical devices, as well as participating in any available and relevant External Quality Assurance (EQA) schemes.

10. Conclusions

The diversity of genomic sequences of SARS-CoV-2 necessitates a proactive approach to assay design and monitoring. To this end these recommendations should be used to guide *in silico* assay monitoring, and to ensure that further in-depth monitoring of clinical assays is maintained. Using the available SARS CoV-2 sequences data in this way will provide the clinical community and those tasked with managing the COVID-19 pandemic with the foresight to avoid diagnostic assay failure in the face of viral genome evolution. Together, with manufacturers of screening and diagnostic SARS-CoV-2 assays, we can help to protect the health of patients and the public.

11. References

ISO - ISO/TS 5798:2022 - In vitro diagnostic test systems — Requirements and recommendations for detection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by nucleic acid amplification methods

SARS-CoV-2: genome sequence prevalence and growth rate - GOV.UK (www.gov.uk)

Guidance for manufacturers: diagnostic assurance with SARS-CoV-2 variants in circulation

Target Product Profile: Laboratory-Based SARS-CoV-2 Viral Detection tests