

METADATA last updated: 2026-03-20 RT initial creation file_name: _AI_FDA SARS-CoV-2 Reference Panel Report - Discrepancies, Governance, and EUA Policy Implications.md file_date: 2026-03-20 title: FloodLAMP FDA SARS-CoV-2 Reference Panel Report: Discrepancies, Governance, and EUA Policy Implications category: regulatory subcategory: fda-policy tags: SARS-CoV-2, COVID-19, diagnostics, EUA, FDA, reference-panel, diagnostic-policy source_file_type: md xfile_type: NA gfile_url: https://docs.google.com/document/d/1JsMxZX6REY4FkNHd5KVepJQOAMuvlQPTIWLGaMnP_Ozo xfile_github_download_url: NA pdf_gdrive_url: NA pdf_github_url: NA conversion_input_file_type: NA conversion: NA license: CC BY 4.0 - <https://creativecommons.org/licenses/by/4.0/> tokens: 10410 words: 7711 notes: Created by GPT-5.4 Pro during archive preparation. **NOT HUMAN VERIFIED - MAY CONTAIN ERRORS** Analyzes the FDA SARS-CoV-2 Reference Panel using the uploaded complaint letter, FDA response letter, McFarlane comparative-data compilation, the Blommel peer-reviewed paper, and additional official FDA, GAO, HHS OIG, WHO/NIBSC, and scientific literature on reference materials, benchmarking, harmonization, and EUA policy. summary_short: FloodLAMP FDA SARS-CoV-2 Reference Panel Report: Discrepancies, Governance, and EUA Policy Implications synthesizes the complaint record, FDA's 2023 response, the comparative LoD data, and relevant scientific literature to assess what the reference panel did well, what remained unresolved, and what this episode implies for diagnostic EUA policy. It argues that the core concept of a common benchmark was sound, but that FDA's transparency, discrepancy-resolution process, and follow-through were inadequate.

CONTENT

Prompt (Cleaned)

Please prepare a comprehensive stand-alone report on the FDA SARS-CoV-2 Reference Panel, which was used to characterize molecular diagnostic tests during the COVID-19 pandemic. Review all of the files I uploaded, synthesize them into a single report, and provide the report as a downloadable file.

The report should be thorough, objective, and fair, but it should also be appropriately critical of the FDA's handling of the reference panel. One of the central issues is the large discrepancy between the limits of detection measured with the FDA reference panel and the self-reported limits of detection on which many EUA tests were originally authorized. As far as I know, the FDA never publicly recognized the scale of this discrepancy, did not clearly report on it, did not resolve it, did not continue follow-up rounds of testing, and did not make the reference panel material broadly available. It seems the agency could have worked with a commercial provider to make the material widely available, kept testing, and tried to resolve or at least narrow the discrepancies, but instead appears to have archived the results and moved on.

Please summarize the December 2020 complaint letter from the attorneys and the FDA's response, which appears to have been issued almost three years later. However, that should be only one part of the report. I also want you to search for additional material, especially scientific

papers and articles about the FDA reference panel, summarize the most relevant ones, and provide citations.

A major focus of the report should be the policy implications for the EUA framework and diagnostic-test regulation more broadly. EUA decisions depend heavily on sponsor-generated performance data. If those data diverge substantially from later standardized measurements, that raises questions about how much confidence regulators and users should place in either dataset and exposes a potential structural weakness in the diagnostic regulatory framework. Please analyze those implications carefully and propose concrete policy solutions to address the gaps revealed by this episode.

Use the uploaded prompt-format file to match the desired saved-file format. I also uploaded a peer-reviewed paper that appears especially relevant; please cite and summarize it.

Files included in context window:

- 2023-09-23_FDA Response Letter - To Complaint regarding FDA Reference Panel from Dec 2020.md
- FDA SARS-CoV-2 Reference Panel Comparative Data - Compiled by Matt McFarlane.md
- 2020-12-22_Complaint Letter to FDA regarding Reference Panel - IQA-Request-Hyman-Phelps-McNamara.md
- 2022-12-15_Paper - Blommel - Authorized SARS-CoV-2 molecular methods show wide variability in the limit of detection.md
- prompt_save-ai-response-to-archive-md.md

FDA SARS-CoV-2 Reference Panel: Discrepancies, Governance, and EUA Policy Implications

Executive Summary

The FDA SARS-CoV-2 Reference Panel was created for a legitimate and, in many respects, important reason: the original analytical sensitivity claims in early SARS-CoV-2 Emergency Use Authorization (EUA) submissions were not directly comparable across tests because developers used different materials, matrices, study designs, and units of measurement.[2][4][5][6][14][15] FDA therefore attempted to create a common benchmark.

That basic idea was sound. The problem is that the program's public record reveals a large and still unexplained divergence between many tests' original EUA-reported limits of detection (LoDs) and the values later reported from FDA reference-panel testing. The strongest peer-reviewed paper directly analyzing the issue, Blommel et al., found a statistically significant difference between mean EUA LoD and mean FDA reference-panel LoD and concluded that the reason for the discrepancy remained unexplained.[4] The uploaded McFarlane compilation,

which is a secondary compilation rather than a primary FDA dataset, points in the same direction: among the 118 rows with numeric FDA/EUA ratios, 101 (85.6%) show a worse LoD on the FDA panel than in the original EUA documentation, with a median FDA/EUA ratio of 9.0x and 57/118 (48.3%) showing a discrepancy of at least 10x.[3]

The December 2020 complaint letter filed by Hyman, Phelps & McNamara argued that the problem lay in the panel material and protocol, especially the use of heat-inactivated virus and dilution into negative clinical matrix, and requested removal of the public comparative data, a public corrective statement, and publication of validation data.[1] FDA's September 29, 2023 response - 1,011 days later - denied the complaint, defended the panel as scientifically sound for its intended purpose, argued that the comparative data were not "influential" information under Information Quality Act standards, and said the data had already been removed because they were outdated, not because of accuracy concerns.[2]

A fair assessment is more mixed than either side's strongest claims. The complaint letter likely overstates what has been proved; the public record does not establish that all or even most divergence was caused by a flawed FDA panel rather than by a combination of panel effects, sponsor-study heterogeneity, protocol differences, matrix effects, or execution variability.[1][4][13][17][18] But FDA's response also leaves the central empirical problem unresolved. FDA did not publicly release the pilot study it cited, did not publish a SARS-CoV-2 reference-panel methods paper by the time of its 2023 response, did not publicly document a structured discrepancy-resolution program, and does not appear - in the sources reviewed here - to have publicly reconciled the conflicting performance claims before the data were archived.[2][7]

The most serious policy lesson is not simply that one panel may have been imperfect. It is that the EUA system lacked a clearly defined mechanism for adjudicating conflicts between sponsor-generated preauthorization performance data and later standardized benchmark results. In effect, the system generated two competing bodies of evidence for the same tests and then failed to explain, resolve, or institutionalize the conflict. That is a governance problem, not only a laboratory problem.

The strongest criticism of FDA is therefore narrower - and stronger - than a claim that the panel is proved invalid. FDA had a valid reason to create a common benchmark, but once substantial discrepancies emerged, it did not provide enough public transparency, follow-up, independent adjudication, or material access to turn the reference panel into a durable public standard. That failure matters beyond COVID-19 because future emergencies will raise the same question: what should regulators do when rapid sponsor-submitted validation data and later standardized comparative data diverge?

Method and Source Base

This report synthesizes four uploaded source documents plus additional official and scientific literature. The uploaded files are: the December 22, 2020 complaint letter, FDA's September 29, 2023 response, the McFarlane compilation of comparative LoD data, and the Blommel peer-reviewed paper.[1][2][3][4] I also reviewed official FDA pages and assessments describing the panel's rollout and later lessons learned; GAO and HHS Office of Inspector General (OIG) reports on FDA's COVID-19 test EUA response; and scientific and policy literature on diagnostic benchmarking, reference materials, assay harmonization, and external quality assessment.[5]-[27]

I also performed a descriptive analysis of the uploaded McFarlane compilation.[3] That analysis should be treated as illustrative rather than definitive for three reasons. First, the file is a secondary compilation, not a primary FDA export. Second, some rows contain non-comparable units (for example PFU or TCID₅₀) or missing values, so ratio statistics only use rows with numeric FDA/EUA ratios. Third, some rows note version-specific caveats. Accordingly, the compilation is best used to characterize the direction and scale of the discrepancy problem, not to substitute for a fully reconstructed primary-source regulatory database.

1. What the FDA Reference Panel Was Supposed to Do

FDA's rationale for the SARS-CoV-2 reference panel was straightforward. Early in the pandemic, many assays were authorized using contrived specimens because natural clinical specimens were scarce. Developers used different source materials - live virus, inactivated virus, extracted RNA, synthetic transcripts, or synthetic genomes - and reported LoD in non-standard units. That made direct comparison across assays difficult or impossible.[2][4][14][15]

FDA presented the reference panel as a common, independent benchmark for nucleic acid amplification tests (NAATs). In its May 28, 2020 Daily Roundup, FDA called the panel "an independent performance validation step" and said it was available to commercial and laboratory developers interacting with FDA through the pre-EUA process or holding an EUA.[5] FDA's archived "Regulatory Science Research Tools" page later described the panel as containing common, independent, well-characterized reference material available to developers of SARS-CoV-2 NAATs for which EUA was requested.[6]

The broader concept was not novel. FDA had used an analogous approach during the Zika emergency, and the Zika reference panel was described in peer-reviewed literature.[2][11] The general scientific case for standardized reference materials is also strong. Several papers from the COVID-19 period argued that assay comparison, inter-laboratory harmonization, and viral-load interpretation all require common standards or reference materials.[12][14][17][18][19][20][21][22]

2. Timeline

Date	Event	Notes
January 2020	CDRH and CBER began collaborating on a molecular reference panel for SARS-CoV-2	Reported in Booz Allen assessment.[7]
February 2020	FDA obtained viral material / live virus for the panel	Booz Allen and FDA response both describe viral material becoming available in February 2020.[2][7]
May 27-28, 2020	Reference panels and protocol became available; FDA publicly announced the panel	FDA described it as an "independent performance validation step." [5][7]
September 15, 2020	FDA first posted comparative results	Reported in Booz Allen assessment.[7]
December 2020	Comparative data last updated on FDA website; complaint letter filed on December 22	Booz Allen says last update was December 2020; complaint filed December 22, 2020.[1][7]
September 29, 2023	FDA denied the complaint	FDA said the data had already been removed as outdated and the panel had been discontinued when materials effectively expired.[2]
January 2025	FDA issued draft guidance on validation of IVDs for emerging pathogens during a Section 564 emergency	FDA says the draft guidance responds in part to Booz Allen and OIG recommendations.[10]

This timeline matters because it shows both the speed and the limit of FDA's follow-through. FDA moved quickly to create a common benchmark in 2020. But the public record reviewed

here does not show an equally robust later phase of discrepancy resolution, republication of validation details, or maintenance of the panel as a continuing public standard.[2][5][6][7][10]

3. What the Public Record Shows About the Discrepancy Problem

3.1 The strongest direct study: Blommel et al.

Blommel et al. analyzed 247 SARS-CoV-2 molecular tests that had received EUA and directly compared LoD values reported in EUA documentation with those reported through the FDA reference panel for the subset of participating tests.[4] Their key findings are the most important single scientific datapoint in this record:

- 130 of 206 developers contacted by FDA participated in the reference-panel challenge.
- Mean EUA LoD was 9,417 copies/mL, while mean FDA reference-panel LoD was 43,750 copies/mL.
- The difference was statistically significant ($p < 0.0001$).
- Within-test comparison also showed the FDA reference-panel LoD was higher in almost all cases, again with $p < 0.0001$.
- None of the variables they tested - method, number of targets, material type, sample type, or control type - explained the discrepancy.
- The authors concluded that the reason for the difference remained unexplained.[4]

This paper is important because it does not merely note wide cross-test variability. It identifies systematic intra-test divergence between the values on which tests were authorized and the later values generated by standardized FDA panel testing. At the same time, the paper is careful: it does not claim to have proved the panel invalid. It explicitly states that the reason for the difference remains unexplained.[4]

Blommel et al. also noted that one manufacturer stated in EUA documentation that the matrix used in the FDA reference-panel study - Minimal Essential Media - had not been evaluated in its interfering-substances study and might reduce sensitivity.[4] That does not prove the complaint's theory, but it is a concrete signal that matrix or commutability effects were at least plausible and not purely speculative.

3.2 Descriptive analysis of the uploaded comparative-data compilation

The uploaded McFarlane file compiles FDA-panel LoDs, EUA LoDs, and FDA/EUA ratios for a large set of assays.[3] Using only rows with numeric FDA/EUA ratios, I obtained the following descriptive summary.

Metric	Value	Interpretation
Rows with numeric FDA/EUA ratios	118	Rows with non-numeric or missing ratios were excluded.
FDA panel LoD > EUA LoD	101 / 118 (85.6%)	In most comparable rows, FDA panel testing suggested worse sensitivity than originally reported.
FDA panel LoD < EUA LoD	17 / 118 (14.4%)	A minority of tests looked better on the FDA panel than in their original EUA documentation.
Median FDA/EUA ratio	9.0x	A typical comparable test had an FDA panel LoD about nine times higher than its EUA LoD.
Geometric mean FDA/EUA ratio	10.1x	Indicates systematic right-skew toward worse FDA-panel LoDs.
Interquartile range	2.0x to 44.6x	Discrepancies were often not marginal.
Ratio >= 10x	57 / 118 (48.3%)	Nearly half of comparable rows were at least ten-fold worse on the FDA panel.
Ratio >= 100x	17 / 118 (14.4%)	Large outliers were not rare.
Ratio >= 1000x	2 / 118 (1.7%)	Extreme outliers existed.

These figures do not prove that FDA was right or wrong about the cause. They do show that the discrepancy problem was too large and too systematic to dismiss as noise.

By release wave, the same compilation suggests that the pattern persisted across multiple data releases rather than being confined to one early batch.

Data release wave	Comparable rows	Median FDA/EUA ratio	Geometric mean	Share with FDA/EUA > 1	Share with FDA/EUA >= 10
1st data release	54	5.3x	4.9x	79.6%	37.0%
2nd data release	20	32.8x	34.3x	100.0%	65.0%
3rd data release	44	14.4x	14.3x	86.4%	54.5%

Again, because this is a secondary compilation, those values should not be treated as definitive regulatory statistics. But they are directionally consistent with Blommel et al. and strongly reinforce the view that the problem was neither isolated nor trivial.[3][4]

3.3 Examples of the largest discrepancies in the uploaded compilation

Test (short name)	EUA LoD	FDA panel LoD	FDA/EUA ratio	Release wave
Diagnostic Solutions	10	54000	5400.0x	3rd data release
MicroGen DX	300	540000	1800.0x	2nd data release
LCT	609	540000	886.7x	3rd data release
Thermo Fisher	750	540000	720.0x	2nd data release
GenMark (Panel)	750	540000	720.0x	3rd data release
Boston Medical	3000	1800000	600.0x	2nd data release
Luminex (Aries)	999	540000	540.5x	1st data release
Roche (infA/B+CoV2, Liat)	36	16200	450.0x	3rd data release
SDI	1500	540000	360.0x	2nd data release

Test (short name)	EUA LoD	FDA panel LoD	FDA/EUA ratio	Release wave
Qiagen	1500	540000	360.0x	1st data release

At the other end, some tests appeared more sensitive on the FDA panel than in their original EUA documentation:

Test (short name)	EUA LoD	FDA panel LoD	FDA/EUA ratio	Release wave
Patients Choice	300000	16200	0.054x	1st data release
LabGenomics	60000	5400	0.090x	1st data release
ScienCell	9600	1620	0.169x	1st data release
Ethos	60000	16200	0.270x	1st data release
CSI Labs	18750	5400	0.288x	3rd data release
Avellino Lab USA	165000	54000	0.327x	1st data release
UMass	12000	5400	0.450x	3rd data release
Gencurix	18000	8100	0.450x	3rd data release

That asymmetry matters. The panel was not simply producing uniformly worse results for everyone. But the overall pattern remained strongly skewed toward worse FDA-panel LoDs, which is why FDA's public response arguably needed more than a general defense of the program.

3.4 Other literature signals consistent with a real discrepancy problem

The broader literature contains several additional signals that should not be ignored.

Arnaout et al. argued that LoD meaningfully affects clinical performance and estimated that each 10-fold increase in LoD reduces sensitivity by about 13%.[12] They explicitly argued that assays should be benchmarked against a universal standard. Their paper does not adjudicate

the FDA panel dispute directly, but it explains why large analytical-sensitivity discrepancies are not just regulatory bookkeeping.

Hirschhorn et al. reported that their internally verified LoD of 100 copies/mL for the Abbott RealTime assay contrasted sharply with the 5,400 NDU/mL value FDA reported using the FDA reference panel.[13] This is one of the clearest published examples of a specific assay whose internally verified analytical performance did not align with the FDA panel result.

Review articles also incorporated the FDA panel into broader discussions of SARS-CoV-2 assay comparison. Yu et al. described the panel as an attempt to solve the problem that IFU/EUA LoDs were not readily comparable because of non-standardized materials and units.[15] Campbell and Binnicker later noted that comparator assays should ideally use an internationally recognized standard or the FDA reference panel to establish sensitivity.[14] Zimmerman et al. noted that many commonly used assays had been compared against an FDA reference panel.[16] Together, these papers show that the panel was not a trivial or purely internal agency exercise; it became part of the diagnostic literature and practice discussion.[14]-[16]

4. What the Complaint Letter Alleged

The December 22, 2020 complaint letter, filed under the Information Quality Act by Hyman, Phelps & McNamara on behalf of unnamed clients, made four main claims.[1]

First, it argued that FDA's public claim that the reference panel allowed more precise comparison across assays was false or misleading because the panel and protocol themselves generated spurious results.[1]

Second, it claimed that the underlying scientific problem was not poor assay performance but flaws in the panel material and protocol. The complaint focused especially on the use of heat-inactivated virus, the assumption that the material was a valid surrogate for actual clinical specimens, and the protocol's instruction to premix panel material with negative clinical matrix. According to the letter, several laboratories had observed variable performance of the reference material in the negative matrix and believed that the range-finding protocol could lock assays into an artificially depressed performance range.[1]

Third, the complaint argued that FDA had continued to encourage reliance on the comparative data even after being notified of significant problems. The letter cited an FDA antigen template encouraging developers to review the reference-panel results when selecting comparator methods.[1][25] FDA's molecular template likewise encouraged developers to review the reference-panel results when selecting a comparator assay.[26]

Fourth, the complaint argued that the comparative data were "influential scientific [and] statistical information" subject to heightened transparency and reproducibility standards under

the IQA, and that FDA had not provided enough information about how the panel was prepared or validated for qualified outsiders to reproduce or evaluate the results.[1]

The requested remedies were specific: remove the comparative data pending correction, publicly state that publication was being suspended because of accuracy concerns, and make validation data demonstrating the validity of the panel and protocol public.[1]

The complaint is strongest when it points to the existence of a large, consequential discrepancy and the absence of public validation detail. It is weaker when it attributes the discrepancy entirely to flaws in the FDA panel. The public record does not fully substantiate that stronger causal claim.

5. How FDA Responded

FDA's September 29, 2023 response rejected the complaint in substance, while also saying the request to remove the data had become moot because the comparative data had already been taken down as outdated.[2]

The response made five principal arguments.

1. The panel addressed a real comparability problem.

FDA argued that the original EUA LoDs were not directly comparable because developers used different materials and methods, especially early in the pandemic when clinical specimens were scarce.[2]

2. The panel and protocol were scientifically sound for their intended use.

FDA said it developed the panel using the same scientific approach used for the Zika FDA reference panel; that it ran a pilot study with several commercial manufacturers and laboratories before broader distribution; and that the observed spread of results supported, rather than undermined, the panel's fitness for purpose.[2]

3. Participation and public posting were built into the EUA framework.

FDA pointed to Conditions of Authorization requiring developers to evaluate analytical LoD and traceability with FDA-recommended reference materials when available and to update labeling after FDA review.[2][7]

4. The comparative data were not "influential" information for IQA purposes.

FDA argued that the reference-panel data were only one factor among many that providers and laboratories might consider and said it was not aware of evidence that users relied heavily on those data alone.[2]

5. The data were removed because they were outdated, not because FDA accepted the complaint's accuracy concerns.

FDA said it had stopped using the reference panel when the materials effectively reached expiration and had removed the website data as part of routine updating of COVID-related information.[2]

FDA's response is persuasive on one limited but important point: the agency had a legitimate reason to want a common benchmark. It is also correct that standardized comparative data do not collapse into a full measure of clinical performance. But the response is noticeably less satisfying on the central empirical question: if many tests' later standardized LoDs differed substantially from their earlier sponsor-reported LoDs, what exactly caused that divergence, and what did FDA do to resolve it?

FDA's answer was essentially that the panel was scientifically sound and the complainants had not proved otherwise.[2] That is a legal and procedural answer. It is not a full scientific reconciliation.

6. A Fair but Critical Assessment

6.1 What the complaint gets right

The complaint is correct that the divergence between many EUA LoDs and many FDA-panel LoDs was large enough to matter.[1][3][4] It is also correct that FDA was using the comparative data in a way that had real downstream relevance. FDA's own templates encouraged developers to review the reference-panel results when selecting comparator tests.[25][26] Even if that does not settle the IQA legal question of whether the data were formally "influential," it undermines any practical suggestion that the comparative data were trivial.

The complaint is also correct on transparency. By September 2023, FDA was still citing an unpublished pilot study and saying it was developing a manuscript on the SARS-CoV-2 reference panel that had not yet been internally cleared.[2] In the sources reviewed for this report, I did not find a publicly released FDA methods paper that resolved the issue. For a program that was publicly ranking or at least comparatively situating authorized tests, that is thin validation disclosure.

Finally, the complaint is directionally right that once such discrepancies emerged, FDA needed a more visible root-cause analysis and adjudication process than the public record shows.[1][2][4][7]

6.2 What the complaint does not prove

The complaint does not prove that the panel was invalid or that the discrepancy "lies not in the EUA assays, but with flaws in FDA's reference panel" in every or even most cases.[1] Several alternative explanations are plausible:

- sponsor-submitted EUA LoD studies may themselves have been inflated or non-comparable because of heterogeneous materials and methods;
- matrix effects or commutability problems may have affected some assays more than others;
- developer-run execution variability could have contributed;
- some assays may genuinely have performed differently under a standardized material than under their original validation material;
- some divergence may reflect unit-conversion or protocol differences rather than simple accuracy failure.[2][4][13]-[18]

The key point is that the public record does not resolve those possibilities.

6.3 What FDA gets right

FDA is right that the original EUA LoDs were not directly comparable and that a common reference material was a sensible regulatory response.[2][4][14][15] FDA is also right that a reference-panel LoD is not the whole story about real-world clinical performance. Analytical LoD matters, but clinical sensitivity also depends on specimen type, collection quality, patient viral dynamics, workflow, and other factors.[2][12][14][16]

FDA is also right that dispersion in results is not what one would expect from a benchmark that was simply useless in the most obvious sense. The panel did not collapse every assay into the same result.[2][3]

6.4 What FDA does not answer

The most important criticism of FDA is that it did not answer the question it most needed to answer.

The public record shows that FDA:

- created and distributed a common panel quickly;
- required developers to use it;
- published comparative results;
- encouraged developers to consult those results;
- later received a formal complaint about major discrepancies;
- defended the panel as sound; and
- eventually removed the data as outdated after the material expired.[1][2][5]-[7][25][26]

What the public record does not show is a fully public, structured discrepancy-resolution process. There is no public FDA paper in the reviewed sources explaining the panel's commutability, matrix effects, reproducibility across laboratories, and reconciliation with divergent sponsor-reported LoDs. There is no public evidence here of repeated rounds of adjudicative testing for outlier assays, no clear public threshold for when divergence would

trigger reinvestigation, and no public, assay-by-assay explanation for the largest discrepancies.[2][7]

FDA's argument that "if the design of the Reference Panel were flawed, it should have resulted in all devices showing high LoD values with no expected dispersion" is also not logically conclusive.[2] A non-commutable or matrix-biased material could, in principle, affect different assays differently. Dispersion alone does not rule that out. That does not prove the complaint's theory, but it does mean FDA's rebuttal was weaker than it appears at first glance.

6.5 The strongest criticism: governance, not merely chemistry

The deeper problem is governance. FDA effectively generated two different performance narratives for many assays: the sponsor-generated preauthorization narrative and the later reference-panel narrative. Both were FDA-mediated. But the agency did not appear to have a robust public process for adjudicating conflicts between them.

That matters because regulators, developers, purchasers, and clinicians need to know which number to trust, when, and why. If a later benchmark is more standardized than the original sponsor study, the regulator must say so plainly and define how contradictory results will be handled. If the later benchmark has limitations that make it unsuitable for overriding sponsor data, the regulator must also say so plainly. In this case, the public record suggests neither happened adequately.

7. Broader Scientific and Policy Literature

The FDA reference panel did attract attention, though usually as part of wider discussions of assay benchmarking, harmonization, or diagnostic regulation rather than as the sole focus of papers.

Citation	Type	Relevance	Main takeaway
Blommel et al. 2023[4]	Original study	Directly analyzes EUA LoDs and FDA panel LoDs	Large, statistically significant intra-test discrepancy remained unexplained.
Arnaout et al. 2021[12]	Original study	Uses FDA verification/reference material and argues for benchmarking	Each 10-fold LoD increase may reduce sensitivity about 13%; universal standards matter.

Citation	Type	Relevance	Main takeaway
Hirschhorn et al. 2021[13]	Original study	Gives a concrete example of assay LoD diverging from FDA panel result	Internal Abbott LoD of 100 copies/mL contrasted with FDA panel value of 5,400 NDU/mL.
Campbell and Binnicker 2022[14]	Review	Discusses how commercial assays should be compared	Comparator assays should ideally use an internationally recognized standard or FDA panel.
Yu et al. 2021[15]	Review	Explains why the panel was created	IFU/EUA LoDs were not directly comparable because of non-standardized materials and units.
Zimmerman et al. 2021[16]	Review	Discusses molecular and antigen test interpretation	Notes that many assays were compared using the FDA panel.
Vierbaum et al. 2022[17]	Original study	Demonstrates value of standardized RNA reference materials	Reference materials help harmonize assay performance and interpret Ct values.
Gavina et al. 2023[18]	Original study	Multisite harmonization work	Cross-platform variability persists without validated reference standards.
Mercer et al. 2022[19]	Perspective/roadmap	Standards policy	Reference standards are foundational across the testing process, not optional extras.

Citation	Type	Relevance	Main takeaway
Sluimer et al. 2024[20]	Original study	Benchmark testing plus EQA model in the Netherlands	A two-step benchmark-and-EQA system can rapidly improve and maintain testing quality at scale.

The literature does not resolve the FDA dispute, but it does place it in a larger frame. Two propositions can both be true at once:

1. Standardized reference materials are essential.
2. Standardized reference materials do not solve governance problems by themselves; they must be commutable enough for the intended use, broadly documented, quality-controlled over time, and linked to a visible discrepancy-resolution process.[12][17]-[21]

8. Policy Implications for the EUA Framework

8.1 EUA depends heavily on sponsor-generated evidence in the period when evidence is weakest

Emergency authorization exists precisely because full evidence is not yet available. That means the system is most dependent on sponsor-generated data when clinical samples, validated comparators, and common standards are least available. FDA, GAO, and OIG all recognize this reality in different ways.[2][7]-[10]

That does not make sponsor-generated data illegitimate. It means the system needs stronger compensating controls after authorization than were clearly built into the SARS-CoV-2 molecular reference-panel process.

8.2 Benchmarking without adjudication can create a "two-track truth"

The reference panel did not simply add information. It created a second track of performance evidence. For many assays, the second track appeared materially worse than the first.[3][4] Without an explicit adjudication pathway, the result is institutional ambiguity.

This is not a narrow COVID-era problem. Any future emerging-pathogen emergency could reproduce the same structure:

- rapid authorization based on heterogeneous sponsor data;
- later benchmarking under a common material;
- substantial divergence;
- no predefined rule for what happens next.

That is why the lesson should be written into future EUA policy before the next emergency, not after it.

8.3 Transparency standards should be higher when FDA itself mediates comparative public rankings

FDA argued that the comparative data were not "influential" information under IQA doctrine.[2] Whatever the legal merits of that argument, the practical reality is that FDA was publishing comparative data and encouraging developers to consult them.[1][25][26] The burden of methodological transparency should therefore have been higher than it was.

At minimum, the public should have had:

- a complete methods description;
- lot characterization and stability data;
- the pilot-study data FDA cited;
- a documented process for handling major discrepancies;
- a persistent archive of superseded results and explanations for removal.[2][7]

8.4 Public access to benchmark materials is not just a logistics question; it is a governance choice

FDA made the panel available to pre-EUA and EUA developers, not as a generally distributed public standard.[5][6] That limitation may have been unavoidable early in 2020. But the public record suggests that the panel was never turned into a durable, broadly accessible standard, and by 2022 FDA was publicly saying that the original panel was not currently available.[2][27]

That matters because wider availability changes the epistemic structure of the problem. If the material is broadly accessible, outside laboratories, external quality assessment programs, proficiency-testing providers, and researchers can test the test of the test. If the material stays within a narrow regulator-developer loop, external scrutiny is much harder.

WHO and NIBSC provide a contrasting model. WHO established an International Standard for SARS-CoV-2 RNA for NAT assays, and NIBSC publicly catalogs the second WHO International Standard (22/252) for purchase and use in assay calibration and harmonization.[21][22] That does not mean FDA could have instantly done the same in spring 2020. It does show that public or quasi-public dissemination models existed and became operational during the pandemic.

8.5 FDA's molecular panel looked less independently adjudicated than its serology program

One of the most revealing contrasts in the record is between FDA's molecular reference-panel program and the serology program. Booz Allen notes that molecular developers ran the reference panel themselves, whereas serology tests were sent to NIH/NCI for an independent government-run evaluation against a standard panel. Booz Allen also notes that the NCI serology evaluation results and underlying validation data were publicly available on FDA's website and openFDA.[7][23][24]

That does not prove serology governance was perfect. It does show that FDA had a more independent and more publicly documented model in another diagnostic domain during the same emergency. The molecular program therefore looks less like the only feasible option and more like one chosen design among alternatives.

8.6 Postauthorization monitoring cannot rely mainly on self-reporting

Booz Allen and GAO both emphasize that EUA allows authorization with less data than traditional pathways, which increases the importance of postmarket monitoring.[7][8] But GAO also noted that FDA's postmarket oversight relies heavily on mandatory and voluntary reporting by developers, user facilities, providers, and consumers.[8] That is useful for detecting egregious failures. It is not a substitute for systematic comparative re-evaluation when benchmark results diverge sharply from authorization data.

9. Policy Recommendations

The following recommendations are designed to address the specific failure mode revealed by the SARS-CoV-2 reference-panel episode.

9.1 Finalize and operationalize a generic emerging-pathogen validation framework before the next emergency

FDA's January 2025 draft guidance on validating IVDs for emerging pathogens is a constructive step and explicitly responds to Booz Allen and OIG recommendations.[10] That effort should be finalized, stress-tested through exercises, and linked to ready-to-use templates for different outbreak phases. The framework should define what changes when an emergency moves from "scarce specimens, few comparators" to "multiple authorized assays, abundant clinical samples, and mature standards."

9.2 Require an early bridging study between sponsor LoD and benchmark LoD

A reference-panel result should not merely sit next to the original EUA LoD. FDA should require a bridging analysis that explicitly compares:

- sponsor validation material,

- standardized benchmark material,
- matrix effects,
- extraction workflow,
- target design, and
- unit conversion method.

This would not eliminate all disagreement, but it would force the source of disagreement into the open.

9.3 Create a formal discrepancy-adjudication protocol

FDA should predefine quantitative thresholds that trigger adjudication - for example, when benchmark LoD differs from sponsor LoD by more than a specified multiple, or by more than one benchmark concentration tier. Triggered adjudication should include:

- repeat testing with a fresh lot of benchmark material,
- testing in at least one independent laboratory,
- review of sponsor study methods and materials,
- a public status label such as preliminary discrepancy, under investigation, resolved, or confirmed revised performance.

The key reform is not the exact threshold. It is the existence of a transparent rule.

9.4 Use independent confirmatory testing, not only developer-run testing

For high-volume or high-impact assays, FDA should move beyond developer-run reference-panel testing and incorporate independent confirmatory testing by government labs, contracted third parties, or designated external quality assessment centers. Booz Allen's contrast with NCI-run serology evaluation suggests this is feasible.[7][23][24]

A mixed model would likely be best:

- developer-run testing for scale and speed,
- independent confirmatory testing for governance and trust.

9.5 Make benchmark materials broadly accessible once initial scarcity passes

FDA should plan from the start to transition benchmark materials from a regulator-controlled distribution model to a broader access model, ideally with:

- published certificates of analysis,
- lot-to-lot traceability,
- stability and expiry data,
- clear intended-use statements,
- distribution through qualified public or commercial reference-material providers.

The point is not only convenience. Wider distribution enables external replication and reduces dependence on a single regulator narrative.

9.6 Use common units and traceability to international standards whenever possible

The literature repeatedly identifies non-standard units as a source of confusion.[4][12][14][15][17]-[22] Future EUA frameworks should push earlier toward common units and traceability to internationally recognized standards or calibrated reference materials. Where a WHO, NIBSC, NIST, or equivalent standard exists, FDA should say explicitly how sponsor-reported performance and benchmark-panel performance map onto that unitage.

9.7 Preserve public comparative datasets with versioning, caveats, and archived explanations

When comparative datasets are taken down or archived, they should not simply disappear into the memory hole of a fast-moving emergency website. FDA should preserve:

- archived datasets,
- dates of updates,
- lot information,
- reasons for removal or supersession,
- known limitations,
- links to successor standards or materials.

If data are removed because they are outdated, that explanation should coexist with, not replace, any unresolved concerns about accuracy or interpretation.[2]

9.8 Build benchmark testing into a larger external quality assessment ecosystem

The Dutch benchmark-plus-EQA model described by Sluimer et al. is highly relevant.[20] The best policy is not a one-time benchmark but a continuum:

- entry benchmark,
- blinded EQA rounds,
- proficiency testing,
- variant challenge panels,
- postmarket discrepancy review.

That approach better fits the reality that assays, workflows, reagents, and circulating variants all change over time.

9.9 Separate authorization claims from comparative benchmark claims in public communication

Public-facing FDA tables should clearly distinguish:

- sponsor-submitted LoD used for authorization,
- benchmark-panel LoD,
- whether the values have been reconciled,
- what the comparative value does and does not imply about clinical sensitivity.

The SARS-CoV-2 reference-panel episode shows that publishing a comparative number without a clear interpretive framework invites confusion.

9.10 Treat large sponsor-benchmark divergence as a regulatory signal in its own right

A large discrepancy should not be viewed merely as an awkward communications issue. It should be treated as a regulatory signal that the evidence base for a test may need re-examination. That does not automatically mean revocation, but it should mean some combination of relabeling, repeat testing, postauthorization study, or public caution note.

10. Bottom-Line Conclusions

The FDA was right to see a problem in the early SARS-CoV-2 EUA landscape: original LoD claims were heterogeneous and hard to compare. Creating a common reference panel was a sensible response.[2][4]-[7][14][15]

But a sound idea is not the same as sound governance. Once the panel generated large and systematic divergences from sponsor-reported EUA LoDs, FDA needed to do more than defend the concept. It needed to publish more methods, conduct and disclose a clearer root-cause analysis, provide a formal path for adjudicating discrepancies, and preserve the resulting evidence in a transparent public record.[1]-[4][7][8]

The available evidence does not prove that the FDA panel was invalid. It does show that the discrepancy problem was real, substantial, and never adequately resolved in public. That is enough to justify serious criticism of FDA's handling of the issue.

The broader policy lesson is straightforward. In a future emergency, a regulator should not have to choose between speed and trust. It should authorize quickly when necessary, but it should also have a prewritten, transparent system for what happens when later standardized evidence does not match the earlier authorization record. The SARS-CoV-2 reference-panel episode is best understood as a warning that benchmarking without adjudication is not enough.

References

- [1] Hyman, Phelps & McNamara. "Information Quality Act Complaint" regarding FDA SARS-CoV-2 Reference Panel Comparative Data, December 22, 2020. Uploaded file: [2020-12-22_Complaint Letter to FDA regarding Reference Panel - IQA-Request-Hyman-Phelps-McNamara.md](#).
- [2] FDA Center for Devices and Radiological Health. Response letter to complaint regarding FDA Reference Panel from December 2020, September 29, 2023. Uploaded file: [2023-09-23_FDA Response Letter - To Complaint regarding FDA Reference Panel from Dec 2020.md](#).
- [3] McFarlane, Matt. "FDA SARS-CoV-2 Reference Panel Comparative Data - Compiled by Matt McFarlane," January 11, 2021. Uploaded file: [FDA SARS-CoV-2 Reference Panel Comparative Data - Compiled by Matt McFarlane.md](#).
- [4] Blommel JH, Jenkinson G, Binnicker MJ, Karon BS, Boccuto L, Ivankovic DS, Sarasua SM, Kipp BR. Authorized SARS-CoV-2 molecular methods show wide variability in the limit of detection. *Diagnostic Microbiology and Infectious Disease*. 2023;105:115880. doi:10.1016/j.diagmicrobio.2022.115880. Uploaded file: [2022-12-15_Paper - Blommel - Authorized SARS-CoV-2 molecular methods show wide variability in the limit of detection.md](#).
- [5] FDA. Coronavirus (COVID-19) Update: Daily Roundup May 28, 2020. FDA website. Announced the SARS-CoV-2 reference panel as an "independent performance validation step" available to commercial and laboratory developers in pre-EUA or with EUA status.
- [6] FDA. Regulatory Science Research Tools. FDA website, content current as of October 1, 2024. Archived description of the FDA SARS-CoV-2 Reference Panel and archived Zika reference materials pages.
- [7] Booz Allen Hamilton. *Deliverable 15: Emergency Use Authorization Assessment - Final Report*. 2021. FDA-posted independent assessment of the agency's COVID-19 EUA response.
- [8] U.S. Government Accountability Office (GAO). *COVID-19: FDA Took Steps to Help Make Tests Available; Policy for Future Public Health Emergencies Needed*. GAO-22-104266. May 2022.
- [9] HHS Office of Inspector General. *FDA Repeatedly Adapted Emergency Use Authorization Policies To Address the Need for COVID-19 Testing*. OEI-01-20-00380. Issued September 16, 2022; posted September 21, 2022.

[10] FDA. *Validation of Certain In Vitro Diagnostic Devices for Emerging Pathogens During a Section 564 Declared Emergency*. Draft Guidance for Industry and Food and Drug Administration Staff. January 2025.

[11] Garcia M, Fares-Gusmao R, Sapsford K, Chancey C, Grinev A, Lovell S, Scherf U, Rios M. A Zika Reference Panel for Molecular-Based Diagnostic Devices as a US Food and Drug Administration Response Tool to a Public Health Emergency. *Journal of Molecular Diagnostics*. 2019;21(6):1025-1033. doi:10.1016/j.jmoldx.2019.06.004.

[12] Arnaout R, Lee RA, Lee GR, Callahan C, Cheng A, Yen CF, Smith KP, Arora R, Kirby JE. The Limit of Detection Matters: The Case for Benchmarking Severe Acute Respiratory Syndrome Coronavirus 2 Testing. *Clinical Infectious Diseases*. 2021;73(9):e3042-e3046. doi:10.1093/cid/ciaa1382.

[13] Hirschhorn JW, Kegl A, Dickerson CZ, et al. Verification and Validation of SARS-CoV-2 Assay Performance on the Abbott m2000 and Alinity m Systems. *Journal of Clinical Microbiology*. 2021;59(5):e03119-20. doi:10.1128/JCM.03119-20.

[14] Campbell MR, Binnicker MJ. Analytic and Clinical Performance of Major Commercial Severe Acute Respiratory Syndrome Coronavirus 2 Molecular Assays in the United States. *Clinics in Laboratory Medicine*. 2022;42(2):129-145. doi:10.1016/j.cll.2022.02.001.

[15] Yu CY, Chan KG, Yean CY, Ang GY. Nucleic Acid-Based Diagnostic Tests for the Detection SARS-CoV-2: An Update. *Diagnostics (Basel)*. 2021;11(1):53. doi:10.3390/diagnostics11010053.

[16] Zimmerman PA, King CL, Ghannoum M, Bonomo RA, Procop GW. Molecular Diagnosis of SARS-CoV-2: Assessing and Interpreting Nucleic Acid and Antigen Tests. *Pathogens and Immunity*. 2021;6(1):135-156. doi:10.20411/pai.v6i1.422.

[17] Vierbaum L, Wojtalewicz N, Grunert HP, et al. RNA reference materials with defined viral RNA loads of SARS-CoV-2 - A useful tool towards a better PCR assay harmonization. *PLoS One*. 2022;17(1):e0262656. doi:10.1371/journal.pone.0262656.

[18] Gavina K, Franco LC, Robinson CM, et al. Standardization of SARS-CoV-2 Cycle Threshold Values: Multisite Investigation Evaluating Viral Quantitation across Multiple Commercial COVID-19 Detection Platforms. *Microbiology Spectrum*. 2023;11(1):e04470-22. doi:10.1128/spectrum.04470-22.

[19] Mercer T, Almond N, Crone MA, et al. The Coronavirus Standards Working Group's roadmap for improved population testing. *Nature Biotechnology*. 2022;40(11):1563-1568. doi:10.1038/s41587-022-01538-1.

[20] Sluimer J, van den Akker WMR, Goderski G, et al. High quality of SARS-CoV-2 molecular diagnostics in a diverse laboratory landscape through supported benchmark testing and External Quality Assessment. *Scientific Reports*. 2024;14:1378. doi:10.1038/s41598-023-50912-9.

[21] World Health Organization. Coronavirus disease (COVID-19) standards and related materials page. WHO notes that the intended use of the International Standard is calibration and control of NAT assays for SARS-CoV-2 RNA.

[22] National Institute for Biological Standards and Control (NIBSC). 2nd WHO International Standard for SARS-CoV-2 RNA (22/252). Public catalog listing showing ongoing product availability for assay calibration/harmonization.

[23] openFDA. Independent Evaluations of COVID-19 Serological Tests. Dataset and API describing government-run independent serology evaluations.

[24] FDA. EUA Authorized Serology Test Performance. FDA website.

[25] FDA. Antigen Template for Test Developers. Official template stating that developers are encouraged to review the FDA SARS-CoV-2 Reference Panel when selecting a comparator method.

[26] FDA. Template for Developers of Molecular Diagnostic Tests for SARS-CoV-2. Official template stating that developers are encouraged to review the FDA SARS-CoV-2 Reference Panel and discuss comparator choice with FDA.

[27] FDA. Coronavirus (COVID-19) Test Development and Validation town hall material, May 2022. FDA noted that the earlier FDA-produced reference panel was not currently available for fulfillment of the condition of authorization unless and until FDA identified current recommended reference materials.