Reviewers' comments:

Reviewer #1: The authors reviewed selected 264 published and pre-print studies on SARS-CoV-2 detection by RT-PCR and virus isolation, aiming to support the laboratory response and contribute to the Public Health control. This reviewer appreciates the authors' aim and effort, and this manuscript can be useful for other researchers, public health officers, and laboratory staff, to find published and pre-print papers necessary for them.
Thank you for the positive review.

However, the current form of manuscript was unsuccessful to be a systematic review. This manuscript rather provided a list of papers with brief description of individual studies. Also, the conclusion obtained with this review was quite general and already made by many previous papers. This reviewer would suggest making or re-organize as a standard and more concise review manuscript citing only highly qualified, verified, and less numbers of selected manuscripts.
We have indeed opted for a rapid instead of systematic review format, aiming to quickly provide a diagnostics snapshot of the first few months of the pandemic.
Thank you for the suggestion of being more selective. However, we felt that providing a dataset of more publications (without selection bias) from the reviewed period was one of the strengths of the paper and could be a helpful "index" for some readers interested in doing their own in-depth data-mining.

Major comments
1. How many preprints were included in the 264 papers? Was inclusion of preprints appropriate?
The review included four preprints. The reason to include them was that they contained key information, not featured in other peer-reviewed publications before our cut-off date, for example comparison between different respiratory samples (Yang et al, medRxiv), or SARS-CoV-2 RNA detection in saliva (Wyllie et al, medRxiv, and Chen et al, SSRN Electronic Journal) and in the male reproductive tract (Ning et al, Preprints).

2. What is 'dpi'? Does it mean 'days post infection'? May the authors intend to describe 'days after onset of illness'?
Thank you for pointing it out. We abbreviated "days post-illness onset" by dpi, introducing it in the Methods section. Indeed it might be confused with "days post infection", thus we replaced it with ...

3. The authors provided with several key points for respiratory shedding. (1) Viral RNA loads peak within the first infection days in the upper and later in the lower respiratory tract. (2) CT findings could precede viral RNA detection in the upper respiratory tract. Are they consist with?
The question of timing the sampling and documenting it well within the disease timeline was something we persistently looked for in the publications we reviewed. Indeed there was evidence for SARS-CoV-2 RNA detection in the upper respiratory tract preceding, simultaneously with and following CT findings. The later fuelled a discussion on the reliability of PCR tests in the first months of the pandemic. Thus we pointed in the text that the full clinical and laboratory presentation needs to be evaluated, and cases with CT findings testing initially negative on respiratory sampling do exist, albeit infrequently:
"*Cases with an epidemiologic link, radiologic findings, and an initial negative result should be monitored further by PCR and evaluated in conjunction with their clinical presentation [10, 22, 35, 38, 41, 55, 56]. The discrepancy between URT and LRT test results has triggered a discussion about the lack of sensitivity of PCR testing.*"
Thank you pointing out that the last summary bullet-point raises questions, we have modified it to read:

"• *CT findings could sometimes precede viral RNA detection in the upper respiratory tract and full clinical presentation should always be evaluated*"

4. At least one paper describes the virus isolation from stools. (Wang W et al. JAMA)

Indeed, we described in the second paragraph of the "Gastrointestinal shedding" section: "*SARS-CoV-2 was isolated from stool sample 15 dpi from a COVID-19 patient with severe pneumonia [271] and from two patients without diarrhoea [77].*" where citation 77 refers to:

Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA. 2020 Mar 11:10.1001/jama.2020.3786.

We have also highlighted it in the first summary bullet-point:

"• *SARS-CoV-2 isolation from feces and RNA detection regardless of gastrointestinal symptoms*"

5. At least one paper describes the virus isolation from urine. (Sun J et al. Emerg Microbes Infect)

Thank you for pointing it out. The letter "Isolation of infectious SARS-CoV-2 from urine of a COVID-19 patient" was published online on 18th May 2020 and thus falls after our cut-off date of 15th May 2020. It is challenging to include all the available information as it becomes outdated on a daily basis.

For completeness, we added a sentence in the revised version mentioning this publication, but have not included it in the study dataset.

"*A letter published shortly after the cut-off date of this review (not included in Supplementary dataset) described successful isolation of SARS-CoV-2 (on Vero E6 cells) at 12 dpi in a severe case [274]..*"

6. Evidence of CNS infection with SARS-CoV-2 is very limited. The authors in this manuscript cited two papers; one detected SARS-CoV-2 RNA in CSF by RT-PCR, but a false-positivity was not excluded in the paper, and the other detected virus-like particle in the brain by EM and viral RNA by RT-PCR. The evidence in the latter paper may be stronger than the former one, but we must be very careful to discuss the CNS involvement by SARS-CoV-2 by only one case in the single paper.

We agree with your comment that the information on COVID-19 and the CNS is limited. In this review we aimed to include all types of samples studied, and mention both detection and lack of it, for a complete overview of the laboratory evidence. Thus we tried to objectively list the available studies (n=8) for CSF, and then concluded in Discussion "*Neurological signs and syndromes associated with COVID-19 and the diagnostic/monitoring potential of CSF testing remain to be clarified.*" We did not discuss whether there is CNS involvement in COVID-19 from the clinical perspective as this is not the scope of the review.

False positives were indeed not excluded in the case report by Moriguchi et al. and the PCR assay used was developed and used primarily in Japan. For additional clarity, we included the available PCR assay details "*(Ct>36 for N target only in a N/N2-based Japanese assay)*" in the revised version of the review.

Minor comments

1. Oral fluid is not the same as saliva.

Indeed, we also had a long discussion on the different types of oral samples and what exactly was sampled in each of the studies we reviewed. Several ones used term "saliva", but collected the sample differently and we felt the term was not always used appropriately. As a consensus, we used oral fluid sampling to generalise on the number of studies: "*Fifteen studies reported on oral fluid sampling (with varying collection methods)...*" and afterwards, where available, described the collection method: "*Self-*

collected deep throat (posterior oropharyngeal) saliva…", "…throat wash with saline solution…". When the collection method was not described in the publication, we used the term as given by the authors, id.e. "saliva".
Furthermore, we now added "saliva" in the key points summary:
"• *Oral fluid/saliva as a self-collectable alternative to respiratory sampling*
• *SARS-CoV-2 RNA detection in oral fluid/saliva up to 4 weeks"*

2. Nasopharyngeal swab may be different from nasal swab.
Indeed and the good laboratory practice suggests collecting nasopharyngeal swabs when testing for viral infections. We tried to check whether that was the case, but information on the sample collection was often not available in the methods sections of the publications we reviewed. Furthermore, we sometimes had the impression that authors collected the sample appropriately (citing US CDC collection guidelines, for example), but used the term "nasal" and "nasopharyngeal" indiscriminately. We decided not to exclude publications based on lack of description of the exact respiratory sample collection technique, as that would have eliminated too many publications, incl. key papers from well-known journals too.

3. The resolution of Figure was too low.
Thank you for the comment, we will provide it in alternative format (pdf) in the revised version.

Reviewer #2:
Suggest some commentary about pre-analytic issues - swab types (flocked, non-flocked, different materials etc), different viral transport media (commercial ones vary eg may contain RNAses if BSA used), limited studies on these pre-analytic issues.
Thank you for directing the discussion to pre-analytical issues. Indeed, we also did not find enough studies discussing them, although we know from experience that laboratories encounter them routinely. We added a sentence in Discussion touching upon the issue:
"*Full comparability of the assays could not be assured due to different factors like materials used (swab types, viral transport media, in-house vs. commercial kits), laboratory equipment, etc. that were infrequently detailed in the publications."*

Authors don't mention saliva as a swab type; role of self-collection versus healthcare worker-performed sample collection.
We had a long discussion on the different types of oral samples and what exactly was sampled in each of the studies we reviewed. Where available, we described the collection method: "*Self-collected deep throat (posterior oropharyngeal) saliva…"*, "…throat wash with saline solution…". When the collection method was not described in the publication, we used the term as given by the authors, id.e. "saliva". We can only assume whether some of the saliva samples were collected by swabbing. Unfortunately, we could not find a study comparing self-collected vs. healthcare worker-performed sample collection, published in the review capture period.
We added a sentence to elaborate on that gap in the revised version:
"*No study compared the different collection methods, for example, self-collection, sampling by a healthcare worker, swabbing, stimulated secretion, etc."*

Could discuss assessment of testing sensitivity and specificity (NPV, PPV etc) in the context of screening, especially if low disease prevalence in the community ie. are there assay performance issues in screening low risk populations versus known COVID-19 positive cases?
Thank you for the suggestion. We had limited the review to the first months of the pandemic (publications cut-off date 15th May 2020) and in that period data was still being accumulated to aid wide population screening programs. We feel it might go beyond the scope of our review to assess testing strategies in different populations,

though it is undoubtedly an engaging topic for another study. Likewise we had to choose to leave other aspects of importance to the laboratory response out.

Role of culture discussed in persistently positive patients, but need to point out such associations will vary depending on NAT platforms and culture techniques (eg type of Vero cells etc) used for comparisons.
Thank for the comment, indeed we did not initially include culturing details. We now added details on cell lines used for isolation (if that information was available) in the revised version:
"Virus isolation was successful from NP swabs 2 dpi in 2 mild cases (using Vero E6 cells) [120]; NP and OP swabs 4 dpi (using Vero CCL-81 cells) [76] in a mild case [63], and also at 4 dpi from NP swab and nasopharyngeal aspirate in another mild case (using Vero E6 cells) [98]."
"La Scola and colleagues cultured 174 NP swabs and 9 sputum samples testing positive via PCR (from 155 patients total) and succeeded with virus isolation from 129 samples (124 with observable cytopathic effect on Vero E6 cells)."
"An Indian study was successful in isolating SARS-CoV-2 (using Vero CCL-81 cells) from respiratory samples in 9 of 12 samples with Ct values ranging 16-25.1 [217]"
"SARS-CoV-2 was isolated from stool sample 15 dpi from a COVID-19 patient with severe pneumonia (using Vero cells) [271] and from two patients without diarrhoea [77]." Where the first study did not specify the type of Vero cells used and the second did not provide any culturing details.
"In a study involving nine mild cases, virus isolation (on Vero E6 cells) was unsuccessful in stool samples 6-12 dpi from 4 patients, and…"
"A letter published shortly after the cut-off date of this review (not included in Supplementary dataset) described successful isolation of SARS-CoV-2 (on Vero E6 cells) at 12 dpi in a severe case [274]."

We also added a mention of cell lines and laboratory equipment used when discussing limitations:
"Full comparability of the assays could not be assured due to different factors like materials used (swab types, viral transport media, cell lines for culturing, in-house vs. commercial kits), laboratory equipment, etc. that were infrequently detailed in the publications."

Although mentioned on p29, it is a shame the authors do not mention assay performance with different viral gene target primers: assays vary in sensitivity (Ct values) when different targets are used. Commercial and in-house assays may use one, tow, three or more gene targets - this may affect sensitivity. Rapid versus high throughput assays may have different sensitivity. Labs should have mechanisms for assay performance evaluations to monitor for false positives (and false negatives); quality assurance programs are important.
We entirely agree with the reviewer on the points raised. Unfortunately, publications usually did not provide information on the sensitivity/specificity of the assays used, nor how the laboratories performed their quality assurance checks and whether adequate EQAs were in place. We now clarify that as part of the limitations:
"However, relatively few of the reviewed articles provided such detailed data, and we could not assess the data quality."
We also added a sentence in Discussion:
"Regular monitoring of the assays performance, detailed reporting and strict quality assurance mechanisms are vital to molecular diagnostics of SARS-CoV-2.Laboratories involved in SARS-CoV-2 diagnostics should ensure compliance with accreditation schemes with a diagnostic scope (e.g. ISO15189 "Medical laboratories – requirements for quality and competence").

Demands for testing have meant the importation and use of unregulated (eg by FDA/TGA etc) swabs/VTM/extraction and testing kits, which may affect performance.
Indeed, we tried to touch upon these issues with an added sentence in the revised Discussion:
"Full comparability of the assays could not be assured due to different factors like materials used (swab types, viral transport media, in-house vs. commercial kits), laboratory equipment, etc. that were infrequently detailed in the publications."

Some mention of the role of serology in confirming uncertain results eg false positives, persistent high Ct samples etc would round out the discussion.
We have consciously decided to omit serology from the manuscript as the manuscript turned out to be too lengthy and a focus on serology could have filled a manuscript by itself. However, we agree it is good to at least acknowledge the supportive role that serology can have in individual patient diagnostics. This besides its important role in studies assessing SARS-CoV-2 antibody prevalence in(sub)populations, characterization of the functionality of immune responses etc.
Therefore we have added the following sentence to the discussion: Although molecular detection on upper respiratory tract samples is the recommended method to diagnose a SARS-CoV-2 infection, serology is occasionally imperative to complement RT-PCR findings as a lack of clinical sensitivity can be observed for RT-PCR-based diagnostics in patients with a strong clinical suspicion for COVID-19. In addition serology can aid in decision making on clinical and infection prevention management, for instance when consistently very low viral loads (high Ct values) are detected in upper respiratory samples. In this case affirmation of the presence of SARS-COV-2 specific (neutralizing) antibodies will inform case management (refs)

Editorial comments (associate editor)

I carefully looked at the paper. It represents a huge amount of information that will be extremely useful for many researchers on COVID-19.
Thank you for the positive evaluation.

I do agree with most of the remarks of the two reviewers that could help improve the manuscript.
We thank the reviewers for their comments and suggestions. We have tried to incorporate them in the revised version of the manuscript.

Indeed, the format of the review is a bit unusual but this choice of a rapid style is explained by the authors.
As explained, we opted for the rapid style and provided an extended dataset that could serve for further in-depth analyses.

My suggestion to make it a bit more "digestable" would be to include more summary tables in the main text recapitulating the key elements. For instance one for respiratory samples, one for GIT samples and one for other samples.

Also, the discussion would warrant to be more in depth taking into account the reviewers remarks
We have added to the discussion as suggested by different reviewers' comments (detailed above).

Also some discussion about potentially different requirements for diagnostics of symptomatic versus pre-/a-symptomatics for which the clinical context nor timing cannot be taken into account for interpretation of the results ;
Thank you for suggesting more detailed discussion. We have already tried to touch upon the topic in the discussion, while referring to the evidence available from the reviewed publications.
"*Further reports with well-documented sampling time points, comparing the different types of respiratory samples and their diagnostic window of use, are needed. Viral RNA concentrations in the URT peak in the early infection days, including in asymptomatic and mildly symptomatic cases. SARS-CoV-2 was successfully isolated from respiratory samples with data suggesting independent replication potential in both the upper and lower respiratory tract.*"
And now added:
"*Thus, respiratory sampling is the optimal strategy for both symptomatic and pre-/a-symptomatic cases and for the latter should be examined in conjunction with epidemiological evidence and clinical follow-up.*"

in that context some some element of discussion about false positive results especially for low viral load samples would be useful.
Indeed, detailed laboratory data that would allows us to assess the proportion of false-positive and false-negatives was often lacking in the reviewed papers from the first months of the pandemic. We added to the limitations discussion (also see responses to reviewer's comments above).

General editorial

- Research, surveillance and review articles should have structured abstracts (max 250 words); other regular articles should have non-structured abstracts (max 200 words).
- Tables must be created in Word. The full table (title, table, notes) should be inserted in the manuscript directly after the first paragraph in which it is mentioned. As tables must be editable, images are not acceptable. To aid readability in both the online and .pdf versions of the article, portrait-oriented tables are preferred whenever possible (https://www.eurosurveillance.org/for-authors).
- If you present numbers with percentages in Tables, the percentages need to be in a Table column separate from the numbers. When the sample size is small (less than 60), we would not generally give percentages as they are subject to disproportional change with increasing or decreasing numerator and static denominator. The tables should not have any empty cells as design element or because information is not available (NA can be used for example).
- Figures must be provided in an editable format, i.e. we need to be able to edit text inside the figure (see our instructions for authors: http://www.eurosurveillance.org/for-authors).
- The supplement files should be headed with a short descriptive title and contain the requested disclaimer at the top (https://www.eurosurveillance.org/for-authors).