

OPEN PEER REVIEW OF THE MANUSCRIPT:

AN ANALYSIS OF SARS-COV-2 VIRAL LOAD BY PATIENT AGE BY TERRY C. JONES, BARBARA MÜHLEMANN, TALITHA VEITH, MARTA ZUCHOWSKI, JÖRG HOFMANN, ANGELA STEIN, ANKE EDELMANN, VICTOR MAX CORMAN, AND CHRISTIAN DROSTEN

BY DOMINIK LIEBL

University of Bonn

The manuscript [Jones et al. \(2020\)](#) was published as a pre-print at the website of the Research Network of Zoonotic Infectious Diseases of the Charité in Berlin, Germany on April 29, 2020 (short link: <https://t.co/xunzyHEi47?amp=1>). The authors communicated their work via twitter* and a national radio show† and received significant international media attention for their central conclusion that “Children may be as infectious as adults.” However, the authors’ very own statistical analysis contradicts their central conclusion. The authors present omnibus test results showing that there exist significant location shifts in the virus load distributions across the age-groups of both age-categorizations. The subsequent post hoc test results demonstrate that the age-group “Kindergarten” (0-6 years) has a significantly lower virus loads in the throat than the age-group “Mature” (> 45 years). The mean viral load of the age-group Kindergarten is by 86% lower than the mean viral load of the age-group Mature.

1. Summary. The authors’ central conclusion – children may be as infectious as adults – is based on the following statistical conclusion:

“Thus the overwhelming conclusion from the three post hoc testing methods is that no significant differences in viral load exist between any subgroups in either categorization.” (See first paragraph on page 4 in [Jones et al. \(2020\)](#))

However, the authors’ very own statistical analysis conflicts with this conclusion since the authors present

- a) omnibus test results showing that there exist significant location shifts in the virus load distributions across the age-groups of both age-categorizations.
- b) post hoc test results that allow to further single out that the age-group “Kindergarten” has a significantly lower virus loads than the age-group “Mature”.

In the following section, I review the authors’ statistical analysis. In Section 3, I comment on better ways to present the statistical results. I will not comment on any medical interpretations of the results since this is not my field of expertise in statistics.

*Link to the tweet: https://twitter.com/c_drosten/status/1255555995671150597?s=20

†Link to the radio show:
<https://www.ndr.de/nachrichten/info/36-Die-Rolle-von-Kindern-ist-nicht-geklaert,audio674236.html>

2. Review of the statistical analysis.

2.1. *Omnibus testing.* The authors start their statistical analysis using a so-called omnibus tests. This type of test is powerful in detecting a possible violation of the authors' null-hypothesis that all age-groups have the same virus load distributions. Since the data are not normally distributed, the authors use a non-parametric, rank-based omnibus test; namely, the Kruskal-Wallis test – a very good choice.¹ The Kruskal-Wallis test shows the existence of significant location shifts in the virus load distribution across the age-groups in either age-categorization C1 and C2 (p-value= 0.008 and p-value= 0.011). These significant omnibus results conflict with the authors' statistical conclusion.

2.2. *Post hoc testing.* Omnibus tests are generally very powerful tests, but are not designed to single out the group(s) with deviating distribution(s). Therefore, in the hope to single out the groups with deviating distributions, the authors use so-called post hoc tests which are, however, by construction substantially less powerful than omnibus tests. The authors propose the following three post hoc testing procedures:

1. Tukey's (Honestly Significant Difference) HSD test
2. Bonferroni-adjusted pairwise *t*-tests
3. The non-parametric, rank-based post hoc test of [Dunn \(1964\)](#).²

However, the first two post hoc tests are parametric tests which assume normally distributed data and which are based on non-robust comparisons of empirical mean values. The use of these two tests is not justified since – as shown by the authors – the data are not normally distributed.

This leaves, only Dunn's test as a valid non-parametric and robust post hoc test which is actually able to single out a significant difference between the virus load distributions of the age-groups "Kindergarten" and "Mature". There is thus by no means an "overwhelming conclusion from the three post hoc testing methods" as stated in [Jones et al. \(2020\)](#).

Remark. Post hoc tests have (by construction) a substantially lower power than omnibus tests. Therefore, there may be further undetected group differences in both categorizations C1 and C2. The Kruskal-Wallis indicated there is at least one further group difference in the age-categorization C1 (p-value 0.008).

The partially very large p-values (often equal to 1) in Tables A6a and A6b in [Jones et al. \(2020\)](#) do not constitute any support of the null-hypothesis of equal virus load distributions across the age-groups. P-values do not measure the probability that the studied hypothesis is true ([Wasserstein and Lazar, 2016](#)).

3. Presentation of the results.

3.1. *Unnecessary log-transformation.* The authors transform the original data using a logarithmic transform to base 10. The non-parametric, rank-based tests used in [Jones et al. \(2020\)](#), however, are invariant to such a logarithm transformation. Thus,

¹The null-hypotheses of normality are rejected by the Shapiro test in both age-categorizations C1 and C2 (both p-values < 0.001).

²I hope the authors refer to the test described in [Dunn \(1964\)](#) since there is also another post hoc test attributed to Dunn which, however, is parametric in nature.

the data could also be analyzed on the original scale. This would simplify the interpretation of the significant differences in the virus loads.

3.2. *On the magnitude of the significant difference.* Table 1 shows the means of the two significantly different groups of categorization C2 as reported in Table 2 of Jones et al. (2020), but back-transformed to their original scale (viral copies per mL). The mean viral load of the age-group Kindergarten is by 86% lower than the mean viral load of the age-group Mature.³

Mean Viral Load per Age-Group	
Kindergarten	Mature
$10^{4.371295}$	$10^{5.229369}$

TABLE 1

Mean viral load values of the significantly different groups of categorization C2 (see Table 2 in Jones et al. (2020)).

Remark. Since non-parametric, rank-based tests procedures are used to analyse the data, one should present the medians - not the means.

4. About the author. My name is Dominik Liebl and I'm a Professor of Statistics at the University Bonn. I assure that I do not follow any political agenda. I'm not involved in any publications or research projects which may lead to a conflict of interest. I also have no financial conflicts of interests. I cannot evaluate the medical statements in the reviewed manuscript – only the statistical analysis as presented by the authors.

References.

- DUNN, O. J. (1964). Multiple comparisons using rank sums. *Technometrics* **6** 241–252.
- JONES, T., MÜHLEMANN, B., VEITH, T., ZUCHOWSKI, M., HOFMANN, J., STEIN, A., EDELMANN, A., CORMAN, V. M. and DROSTEN, C. (2020). An analysis of SARS-CoV-2 viral load by patient age. *German Research network Zoonotic Infectious Diseases website*: <https://t.co/xunzyHEi47?amp=1>.
- WASSERSTEIN, R. L. and LAZAR, N. A. (2016). The ASA statement on p-Values: Context, process, and purpose. *The American Statistician* **70** 129–133.

DOMINIK LIEBL
INSTITUTE OF FINANCE AND STATISTICS
UNIVERSITY OF BONN
AND
HAUSDORFF CENTER OF MATHEMATICS
UNIVERSITY OF BONN
ADENAUERALLEE 24-26
53113 BONN, GERMANY
E-MAIL: dliebl@uni-bonn.de

³ $86.1348\% = (10^{5.229369} - 10^{4.371295})/10^{5.229369}$