

# Review of: "Quantile regression for identifying latent structures in COVID-19 pandemic – Examples from Nepal"

Boudewijn Roukema<sup>1</sup>

<sup>1</sup> Nicolaus Copernicus University

**Potential competing interests:** No potential competing interests to declare.

This article requires major revision before being resubmitted, since in the current state, it is impossible to assess as a scientific research article. Under conventional journal standards, this article should be considered "rejected, but resubmission after a complete rewrite would be considered".

LINEBREAK Below are major concerns focussing on the abstract; general concerns; and major concerns with the Introduction, Data and Method (sub-)sections. A few examples of incorrect or unjustified scientific wording or use of sources are interspersed with the major concerns, and a few concerns in the Results and Conclusions sections are listed.

LINEBREAK A comment on language is included in the 'General' concerns. The English-language barrier should not be used to dismiss a valid scientific paper. However, the scientific reasoning (suggestions, assertions, implications) have to be clear to the reader. Ambiguous statements are unfalsifiable in the Popperian sense.

LINEBREAK == Abstract ==

LINEBREAK \* Abstract: Aim: It is very difficult to guess the aim of the paper. It appears that the aim is primarily to study the effects of vaccination, and secondarily of early pandemic governmental controls, on the numbers of daily SARS-CoV-2 infections in four countries. What is extremely unclear is whether the aim is to provide fitting functions (models) of the temporal evolution of the SARS-CoV-2 daily infection counts, or to infer causal relations. The former seems to be within attainable with the quantile regression method; the latter is a much more ambitious question.

LINEBREAK \* Abstract: Method: It appears that the method of using quantile regression is argued to be a way of allowing causal inferences to be made despite underreporting. If so, then this needs to be clearly stated. The method needs to be stated in a way that it relates to the aim.

LINEBREAK \* Abstract: "different pattern" presumably means "different phases".

LINEBREAK == General == \* The quality of the English should be improved, but there are two aspects to this: (i) ambiguity in scientific reasoning; (ii) non-standard English that can be understood without affecting the scientific reasoning.

LINEBREAK (i) A scientific research paper must have clarity in every sentence. The scientific reasoning - speculation versus educated guess versus reasonable hypothesis versus implication with an implicit or calculated Bayesian probability

- has to be clearly indicated by the words. For example, the difference between "suggests" and "implies" is fundamental. Interpretations made by the author(s) should be attributed, e.g. "We interpret this as ..."

LINEBREAK (ii) Non-standard English and grammatical errors or style aspects that do not put the reasoning into doubt. This is an open question for an open peer review journal such as Qeios: should or can language editing and proofreading be crowd-sourced? The comments here assume that a proofreader will improve the clarity of the text after a revised version is prepared, with (i) the scientific reasoning clarified.

LINEBREAK \* References: The DOIs in articles with DOIs should be presented as clickable links pointing to <https://oadoi.org/DOI> where DOI is the DOI string; this would optimise open peer review by giving access to readers without paid subscription access. See [https://www.w3schools.com/html/html\\_links.asp](https://www.w3schools.com/html/html_links.asp) for raw html code for creating clickable links ("anchors" with the "a" tag).

LINEBREAK \* References: The preprint bibliometric data of the research papers (ArXiv, MedrXiv, BiorXiv) should be provided, since readers immediately know that these versions are open access. Some papers may lack preprint versions, but in fields such as astronomy and cosmology, posting as preprints on ArXiv is de facto obligatory.

LINEBREAK \* Algebraic notation: the author jumps from variables named as words in the data section to symbolic algebraic notation in the Method section. It would be much clearer to assign conventional variable names (latin or greek) to the parameters studied. This will enable the Method section to be written, and parts of the Results sections that are close to unreadable, such as "If at time t, Total Deaths (t) is very high and Total Infected (t) is also very high," can be rewritten correctly, using an italic font for the variables (such as "t"). Subscripts, e.g. "TIC" for "Total Infected Cases" would be written in roman font to indicate that "TIC" does not mean "T times I times C".

LINEBREAK == Introduction == \* Introduction:

LINEBREAK \*\* "massive flow of COVID-19 infected" - Infections are with SARS-CoV-2, not COVID-19.

LINEBREAK \*\* "COVID-19 testing" - probably "SARS-CoV-2 testing" is the intended meaning. LINEBREAK \*\* Ref [3]. A newspaper article is not a valid source for asserting that daily infections and deaths were underreported.

LINEBREAK \*\* A full stop is missing from the end of paragraph 3.

LINEBREAK \*\* Refs [4], [5]. The BBC is not a valid source for the claims attributed to [4] and [5].

LINEBREAK \*\* "Germany and the Netherlands have a good health insurance system. Hence underreporting cases are very low here." The first sentence cannot be asserted without a source. A sentence that might be acceptable without a source could be "Germany and the Netherlands have the reputation of having good health insurance systems."

LINEBREAK The second sentence is not a valid inference from the first. One might *speculate* a causal relation, but assertion is not proof.

LINEBREAK \*\* In the following sentence, "this region" is too vague; the Table in Ref [6] <https://oadoi.org/10.1016/S0140->

6736(21)02796-3 gives a ratio of "1.82 (1.73 to 1.88)" for Germany and "2.13 (2.00 to 2.24)" for Netherlands, so "1 and <2" disagrees with the source if we interpret the text literally, i.e. for "the region of Germany and the Netherlands". If the author wishes to say "Western Europe", then that must be stated.

LINEBREAK \*\* The end of the same paragraph makes the error of writing "COVID-19 testing"; while diagnoses determining COVID-19 exist, this is unlikely to be the intended meaning. Moreover, the whole sentence "Sufficient ... and deaths" is too vague. Underreporting by a factor of about 1/1.82 in Germany cannot be described as "correct reporting" nor "complete reporting".

LINEBREAK \*\* The paragraph "In this paper, quantile regression ..." should first of all give a brief description of the way in which quantile regression would be expected to provide answers to the question or questions defined as the aim of the paper.

LINEBREAK == Data ==

LINEBREAK \* Ref [12] is too vague. Specific URLs of the data files should be provided, and preferably archived (<https://archive.org> or <https://archive.today>) URLs of the files should be provided too. As of the time of writing this review, <https://covid19.mohp.gov.np> is unavailable. If the source data are presented in the author's own format (e.g. plain text files), then they are likely to qualify under copyright principles for posting under a free licence at Zenodo (<https://zenodo.org>) or another service for long-term archiving of research records.

LINEBREAK \* Ref [13] is again too vague. The comments for [12] apply (except that the WHO website is currently online.)

LINEBREAK == Method ==

LINEBREAK \* The section title should be "Method", not "Methodology", since the author does attempt a general study of methods, but instead presents a particular method.

LINEBREAK \* The DOI for Ref [16] is missing: 10.1177/1471082X1875914

LINEBREAK \* The biggest flaw in the Method section is that after a brief introductory section, a description of the way in which the data is to be analysed and inferences drawn is completely missing. Writing the missing part of the method section will require reasonable choice of variable names for the parameters studied.

LINEBREAK \* It is not obligatory to provide the source code (software environment, analysis software) and specific scripts for analysing and plotting the data, but in the context of open peer review and emerging standards of reproducibility, that would strengthen the paper. Ideally, the full paper should be made in a reproducible format such the Manega framework of shell scripts and make files (Akhlaghi+2021, <https://doi.org/10.1109%2FMCSE.2021.3072860>; Roukema 2021, <https://peerj.com/articles/11856> ).

LINEBREAK

LINEBREAK == Results ==

LINEBREAK \* Three of the sub-plots in Figure 4 have an unexplained diagonal line suggesting that the author didn't thoroughly check the input files. Modifications to the original source files to handle unrealistic values would have to be documented in a place where the reader would expect to find them.

LINEBREAK \* Figures 12 and 13 should be provided in a postscript or other appropriate format so that the text can read clearly and that the curves are free of low-resolution artefacts.

LINEBREAK \* Figure 14 has several diagonal lines whose meaning is extremely difficult to guess from the caption; again, this probably indicates that the author didn't study the data source files for obvious errors.

PS: Apologies for the "LINEBREAK" tags! This is a hack because the Qeios interface does not yet permit plain text entry: end-of-line characters - both ^M and ^J - are ignored in text entry. Trying to read a single paragraph with 36 end-of-line characters abusively removed by the Qeios software would be extremely cumbersome without an appropriate marker. I had to click "enter" 36 times to reconstruct the paragraph structure.