ATTACHMENT for "Concerns about an article from HMS and Boston Children's Hospital"

Here are the reasons why we think the article's conclusion is not a scientific one:

1. The language in the article is inconsistent. The authors made misleading claims in the title and abstract. In the title, the authors claim that "...China indicates early disease activity in the Fall of 2019", and the abstract concludes by "The increase of both signals precedes the documented start of the COVID-19 pandemic in December...". However, in the body text, the authors concluded that "...retrospective analysis **cannot verify** if the increased hospital and search engine volume is related to the SARS-CoV-2 virus…" and "...further research is needed to validate the emergence of SARS-CoV-2". Those statements are not reflected in the title/abstract at all. Unfortunately, the press and social media only referred to the title and abstract. Any reasonable person would be aware of the political implications of claims suggesting COVID-19 started in China several months earlier than reported, and any reasonable researcher would be able to foresee misinterpretation of studies like the authors' in lay media. To avoid such misunderstanding, the authors should have maintained the same level of integrity throughout the whole paper, and even be more cautious about the eye-catching title and abstract.

2. One of the two main points the authors make is based on the historical trend in the popularity of the search term "diarrhea" on a Chinese search engine. The authors used an uncommon term "腹泻的症状" (literally translated as "symptom of diarrhea") while query data on more common terms such as diarrhea "腹泻" (what Google Translate showed when translating "diarrhea") and upset stomach "拉肚子" show a contradicting pattern. Nevertheless, the authors failed to acknowledge these contradicting evidence beyond reason to draw their claims. In addition, given that diarrhea ranks **lowest** on the list of symptoms of COVID-19 on the CDC website, we consider it reasonable to question whether the authors ran multiple tests down the list of potential symptoms (also known as data dredging) until a predetermined conclusion was verified, while it is well known that multiple testing finds spurious correlations.

3. The data samples are not representative. There are more than 300 hospitals in Wuhan, but the authors only picked 6 hospitals and analyzed their parking lots. Furthermore, one sample from such a small study cohort is Hubei Women and Children's Hospital, which focuses on maternal and pediatric health care, and as a matter of fact, does not take any adult patients with respiratory diseases. We wonder if the authors would also draw the conclusion that many children were hit by COVID-19 in Wuhan. Naturally, it leads us to question whether the authors exerted reasonable efforts to maintain a high level of academic rigorousness.

4. Last but not least, this not-peer-reviewed article was made available since 2020-06-08T11:36:28Z according to Harvard Dash, but the news article from ABC News "Satellite data suggests coronavirus may have hit China earlier: Researchers" was published at an even earlier time, on June 8 at 6:04 am. We don't consider it appropriate for the authors to hastily draw media's attention before fully evaluating the potential negative social impacts, and we don't consider it responsible as researchers to advertise apparently controversial and potentially misleading inconclusive research to the general public before receiving peer reviews.

We are open to discussions with the authors or anyone else who has questions about our points. You may find other criticisms of the article in https://www.bbc.com/news/world-asia-china-53005768.