* Main problem #1: Why exclude the most recent 4 weeks? That is the whole point of nowcasting, to use the reporting delay distributions over recent weeks. Simply excluding the last 4 weeks of data seems very counterintuitive, since nowcasting needs to use those weeks to account for reporting delays. This seems like a substantial problem with a nowcasting analysis. Of course the numbers are incomplete, because of the reporting delays.

Answer: Our goal is to test whether a simple linear model plus a real-time proxy (Google Trends) can rival complex delay-correction methods. We found Google Trends and finalized case counts to be highly correlated. However, the last four weeks of reported cases suffer from severe underreporting that simple linear method cannot flexibly adjust for.

By omitting those censored weeks and substituting Trends (which arrive without delay), we preserve model simplicity and directly assess the value added by digital data. Explicit delay modeling, while valuable, is outside this study’s scope and reserved for future work.

* Main problem #2: The authors present an entirely different model (SARIMAX) AND data source and compare it to existing ones. Why not use the existing model from InfoDengue (which has been validated and gone through extensive peer review) and additionally show it with the inclusion of Google Trends data and a combination of the two?

Answer: Our main goal is to test whether Google Trends alone can improve a very simple nowcasting model. We chose linear methods like GT, SARIMAX because it has no built-in delay correction or extra layers. We keep the model basic so any gain clearly comes from the Trends data. If we mix Trends into the full InfoDengue model, we can’t tell if the improvement is due to the new data or the more complex model. Building that combined model goes beyond this proof-of-concept and can be done in future work.

* Main problem #3: It is unclear why the authors chose the performance metrics they did, and why so many. It might be more appropriate to use log scores or weighted interval scoring metrics. Additionally, it might be helpful to also show 50% credible interval coverage. For each of the performance metrics shown, all models (including InfoDengue and the naïve model) should be shown. They are only shown for some of the metrics. It is very important if trying to present a new and improved method/data source to show that is has better performance to the existing or baseline models.

Answer: Do we need log scores/50% credible interval coverage? For log socres, we need to assume normal distribution for example.

Right now, we only show point‐estimate metrics for InfoDengue and the naïve model. InfoDengue does not provide upper credible limits for the most recent weeks in about half of the states. The naïve model simply uses the last observed case count as the forecast for the next time point, so it cannot produce any interval predictions.

* + Can the authors present performance metrics for not only the current week, but the previous week, and the week before that, and the week before that? I.e., at all time horizons of the nowcast. This is standard for nowcasting model performance evaluation.

Answer: 0,1,2,3 weeks, could be done. But for these all together, the result metrics would be too much. We need maybe less metrics.

* + Additionally, are the performance metrics presented for the full timeframe? Dengue incidence and model performance can vary depending on whether it is an epidemic or non-epidemic year. Moreover, the years after Zika was introduced were marked by unusual dengue dynamics. Can the authors show model performance metrics for different years?

Answer: InfoDengue only releases the latest corrected counts: for week T you see N1​, and one week later it’s updated to N1+N2. If you pull data at a single timepoint, you miss all intermediate revisions. To recover the true “real-time” inputs, we have been archiving weekly snapshots of the reported cases since EW 10 2024. This gives us exactly what was known each week—but spans only about one year, so we cannot yet split performance by epidemic vs. non-epidemic years or pre/post-Zika periods.

Moreover, our paper’s aim is a proof-of-concept: show that a very simple model plus Google Trends can compete with more complex nowcasting methods over the available timeframe, rather than perform an exhaustive year-by-year analysis.

* + The models don’t seem to have performed very well, particularly regarding 95% coverage probabilities. Only in a couple regions and models does it reach 95%. Some of them are as low as 0.07. This seems like a substantial problem with the model performance. Also, the width of the uncertainty intervals doesn’t seem like a very helpful metric when the coverage probabilities are so small.

Answer: TBD, we update the result and answer this

**Discussion & Conclusions**

* Some of the conclusions the authors make are based on models that have performed very poorly. In fact, the Google Trends only model seems to have better metrics compared to the other models that are shown, but it isn’t always compared to the naïve model or InfoDengue model, which seem to have decent performance. It is also unclear because of the different model and data whether the difference in model performance is simply because of the new type of data (Google Trends), or whether it has to do with the difference in models. Additionally, the overall performance doesn’t seem strong even with the performance metrics shown.

Answer: Update the result.

The objective of this paper is to use simple linear method to incorporate the cheap digital data, and hopefully beat the complicated model. From the result, the linear model beat the complicated bayesian model in many states. As a simple linear model, it is satisfying. Also, there are some limits of the Infodengue model. It doesn't provide UCL in some states, and it doesn't provide the prediction for most recent week because of the complexity, i.e., multiple data resources required.

%%%%%%%%%%%%%%

\section{Analysis to be run by Yang and Guilherme}

Reviewer 4: Run the analysis with a longer time period

Reviewer 2: Add metrics such as log score and 50\% coverage interval

Reviewer 2: Can the authors present performance metrics for not only the current week, but the previous week, and the week before that, and the week before that? I.e., at all time horizons of the nowcast. This is standard for nowcasting model performance evaluation.

%%%%%%%%%%%%%%