Numbers and data can be critical tools in bringing complex issues into crisp focus. The understanding of diseases, for example, benefits from algorithms that help monitor their spread. But without context, a number may just be a number, or worse, misleading.

In “The Parable of Google Flu: Traps in Big Data Analysis”, the authors examine Google's data-aggregating tool Google Flu Trend (GFT), which was designed to provide real-time monitoring of flu cases around the world based on Google searches that matched terms for flu-related activity.“Google Flu Trend is an amazing piece of engineering and a very useful tool, but it also illustrates where 'big data' analysis can go wrong," said Ryan Kennedy. He and co-researchers David Lazer, Alex Vespignani and Gary King detailed new research about the problematic use of big data from aggregators such as Google.

Even with modifications to the GFT over many years, the tool that set out to improve response to flu outbreaks has overestimated peak flu cases in the U.S. over the past two years."Many sources of 'big data' come from private companies, who, just like Google, are constantly changing their service in accordance with their business model," said Kennedy, who also teaches research methods and statistics for political scientists. "We need a better understanding of how this affects the data they produce; otherwise we run the risk of drawing incorrect conclusions and adopting improper policies."

GFT overestimated the prevalence of flu in the 2012-2013 season, as well as the actual levels of flu in 2011-2012, by more than 50 percent, according to the research. Additionally, from August 2011 to September 2013, GFT over-predicted the prevalence of flu in 100 out of 108 weeks.

The team also questions data collections from platforms such as Twitter and Facebook as campaigns and companies can manipulate these platforms to ensure their products are trending.

Still, the article contends there is room for data from the Googles and Twitters of the Internet to combine with more traditional methodologies, in the name of creating a deeper and more accurate understanding of human behavior.