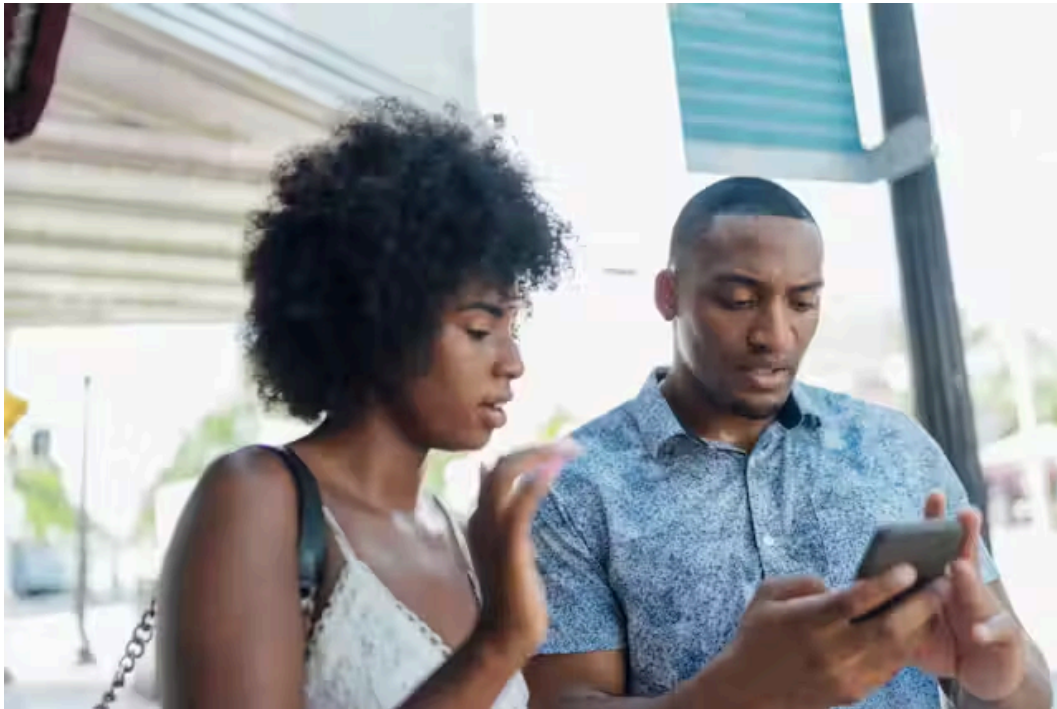


THE CONVERSATION

Academic rigor, journalistic flair



Social media is flush with advice urging non-menstruating people to use period tracking apps in order to trip up the apps' algorithms. Westend61 via Getty Images

No, submitting junk data to period tracking apps won't protect reproductive privacy

Published: July 7, 2022 8:25am EDT

Katie Siek

Professor and Chair of Informatics, Indiana University

Alexander L. Hayes

Ph.D. Student in Health Informatics, Indiana University

Zaidat Ibrahim

Ph.D student in Health Informatics, Indiana University

Social media users posted ideas about how to protect people's reproductive privacy when the Supreme Court overturned *Roe v. Wade*, including entering "junk" data into apps designed for tracking menstrual cycles.

People use period tracking apps to predict their next period, talk to their doctor about their cycle and identify when they are fertile. Users log everything from cravings to period flow, and apps provide predictions based on these inputs. The app predictions help with simple decisions, like when to buy tampons next, and provide life-changing observations, like whether you're pregnant.

The argument for submitting junk data is that doing so will trip up the apps' algorithms, making it difficult or impossible for authorities or vigilantes to use the data to violate people's privacy. That argument, however, doesn't hold water.



As researchers who develop and evaluate technologies that help people manage their health, we analyze how app companies collect data from their users to provide useful services. We know that for popular period tracking applications, millions of people would need to input junk data to even nudge the algorithm.

Also, junk data is a form of “noise,” which is an inherent problem that developers design algorithms to be robust against. Even if junk data successfully “confused” the algorithm or provided too much data for authorities to investigate, the success would be short-lived because the app would be less accurate for its intended purpose and people would stop using it.

In addition, it wouldn't solve existing privacy concerns because people's digital footprints are everywhere, from internet searches to phone app use and location tracking. This is why advice urging people to delete their period tracking apps is well-intentioned but off the mark.

How the apps work

When you first open an app, you input your age, date of your last period, how long your cycle is and what type of birth control you use. Some apps connect to other apps like physical activity trackers. You record relevant information, including when your period starts, cramps, discharge consistency, cravings, sex drive, sexual activity, mood and flow heaviness.

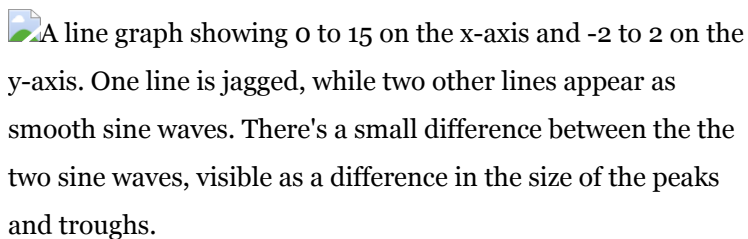
Once you give your data to the period app company, it is unclear exactly what happens to it because the algorithms are proprietary and part of the business model of the company. Some apps ask for the user's cycle length, which people may not know. Indeed, researchers found that 25.3% of people said that their cycle had the oft-cited duration of 28 days; however, only 12.4% actually had a 28-day cycle. So if an app used the data that you input to make predictions about you, it may take a few cycles for the app to calculate your cycle length and more accurately predict the phases of your cycle.

An app could make predictions based on all the data the app company has collected from its users or based on your demographics. For example, the app's algorithm knows that a person with a higher body mass index might have a 36-day cycle. Or it could use a hybrid approach that makes predictions based on your data but compares it with the company's large data set from all its users to let you know what's typical – for example, that a majority of people report having cramps right before their period.

What submitting junk data accomplishes

If you regularly use a period tracking app and give it inaccurate data, the app's personalized predictions, like when your next period will occur, could likewise become inaccurate. If your cycle is 28 days and you start logging that your cycle is now 36 days, the app should adjust – even if that new information is false.

But what about the data in aggregate? The simplest way to combine data from multiple users is to average them. For example, the most popular period tracking app, Flo, has an estimated 230 million users. Imagine three cases: a single user, the average of 230 million users and the average of 230 million users plus 3.5 million users submitting junk data.

A line graph showing 0 to 15 on the x-axis and -2 to 2 on the y-axis. One line is jagged, while two other lines appear as smooth sine waves. There's a small difference between the two sine waves, visible as a difference in the size of the peaks and troughs.

The blue line represents a single user. The orange line is the average of 230 million users. The green line combines 230 million users submitting good data with 3.5 million users submitting junk data. Note that there is little difference between the orange and green lines. Alexander Lee Hayes, CC BY-SA

An individual's data may be noisy, but the underlying trend is more obvious when averaged over many users, smoothing out the noise to make the trend more obvious. Junk data is just another type of noise. The difference between the clean and fouled data is noticeable, but the overall trend in the data is still obvious.

This simple example illustrates three problems. People who submit junk data are unlikely to affect predictions for any individual app user. It would take an extraordinary amount of work to shift the underlying signal across the whole population. And even if this occurred, poisoning the data risks making the app useless for those who need it.

Other approaches to protecting privacy

In response to people's concerns about their period app data being used against them, some period apps made public statements about creating an anonymous mode, using end-to-end encryption and following European privacy laws.

The security of any "anonymous mode" hinges on what it actually does. Flo's statement says that the company will de-identify data by removing names, email addresses and technical identifiers.

Removing names and email addresses is a good start, but the company doesn't define what they mean by technical identifiers.

With Texas paving the road to legally sue anyone aiding anyone else seeking an abortion, and 87% of people in the U.S. identifiable by minimal demographic information like ZIP code, gender and date of birth, any demographic data or identifier has the potential to harm people seeking reproductive health care. There is a massive market for user data, primarily for targeted advertising, that makes it possible to learn a frightening amount about nearly anyone in the U.S.

While end-to-end encryption and the European General Data Protection Regulation (GDPR) can protect your data from legal inquiries, unfortunately none of these solutions help with the digital footprints everyone leaves behind with everyday use of technology. Even users' search histories can identify how far along they are in pregnancy.

What do we really need?

Instead of brainstorming ways to circumvent technology to decrease potential harm and legal trouble, we believe that people should advocate for digital privacy protections and restrictions of data usage and sharing. Companies should effectively communicate and receive feedback from people about how their data is being used, their risk level for exposure to potential harm, and the value of their data to the company.

People have been concerned about digital data collection in recent years. However, in a post-Roe world, more people can be placed at legal risk for doing standard health tracking.