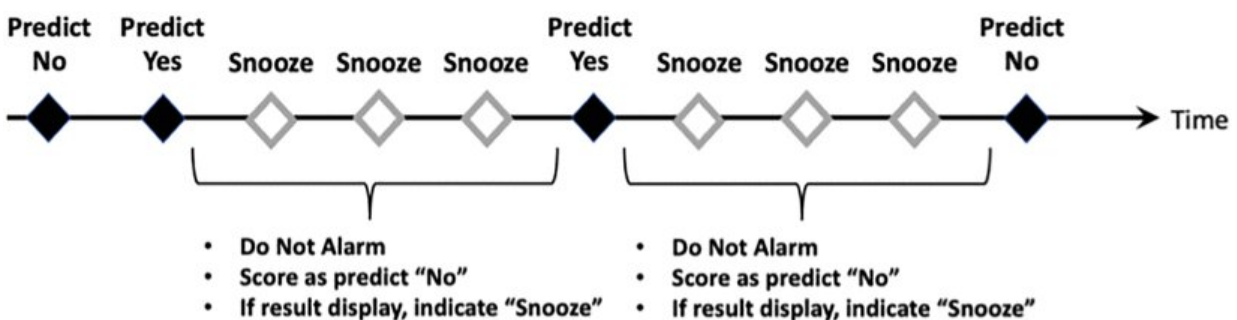# Common medical statistics often wrong or misleading: Study offers a corrective approach

December 12 2022, by Colleen Reynolds

Snoozing acts to suppress any alarms that would be triggered by predictions during the snooze period. This improves u-metrics that incorporate adverse positive utility, such as u-precision, u-specificity, and the adversity ratio. Credit: *IEEE Journal of Biomedical and Health Informatics* (2022). DOI: 10.1109/JBHI.2022.3189312

A simple, yet revolutionary new statistical technique enables better assessment and implementation of many tests and predictive models, leading to greater patient benefits.

Faulty assumptions in some widely used statistics can lead to flawed predictive model implementations that impact patient care. This can be corrected with a novel, utility-based approach ("u-metrics"), according to a study in the *IEEE Journal of Biomedical and Health Informatics*, co-

authored by Dr. Jonathan Handler, Senior Fellow for Innovation at OSF Healthcare.

What's wrong with the classic statistics?

Predictors of whether something will or will not happen in the future are used to facilitate care. Classic statistics to assess these predictors and guide their implementations include sensitivity, specificity, and positive and negative predictive values. These are based only on counts of how often the predictor was right or wrong. The article notes that these classic statistics make assumptions that don't apply to many (probably most) real-world scenarios.

Statistics based on faulty assumptions may suggest that a predictor will yield great benefit to patients even though the real-world performance will prove disappointing or even harmful. The result? Too often, busy health care workers must suffer through frequent false or useless alarms that they soon learn to ignore ("alert fatigue").

For example, a prediction system might incorrectly trigger an alarm, claiming that a patient who is healthy has a dangerous infection. It may also correctly trigger an alarm for a patient with a dangerous infection even though the team is already addressing the issue. In each case, the alarm adds no value and distracts the care team away from other important work.

Worse, in the case of a correct but unhelpful and distracting alarm, classic statistics inappropriately "take credit" for a correct prediction even though the alarm created more harm than benefit. This is because classic statistics assume that correct predictions are always helpful and every correct prediction is equally helpful, even though, as the authors note, those assumptions are commonly not the case.

A new and better approach

To address these challenges, the authors created u-metrics, an intuitive and comprehensive solution that does not rely on assumptions that rarely apply in the real world. Unlike classic statistics, it does not take a one-size-fits-all approach. Instead, it assigns to each prediction only the credit it deserves, and categorizes each prediction based on the benefit or harm created rather than its correctness.

"Health care providers often complain that research suggests predictors will perform well, but when implemented in the real world, the impact is disappointing and sometimes harmful," said Dr. Handler, lead author of the study.

"There has been some limited acknowledgement that the assumptions of classic count-based statistics commonly do not apply and a few partial fixes have been proposed. However, to our knowledge, this is the first comprehensive evaluation of the assumptions required by these classic statistics, and more importantly, the first comprehensive fix to the problem. We believe that using u-metrics to guide the development, selection, and implementation of these types of predictors will benefit patients, providers, and health systems."

In addition to clinical predictors, u-metrics can be used to assess any system that provides yes or no responses, from weather alerts to stock market predictions.

How can u-metrics help address alert fatigue?

By better informing the selection of predictors and their implementations, u-metrics may reduce the likelihood that a health system will choose and operationalize a predictor that fires too many

useless alerts. The paper also describes "snoozing," an implementation technique to dramatically reduce false and nuisance alerts in many cases. Snoozing is when an alarm is automatically or manually silenced for a period of time after it fires.

Although manual snoozing of sensor alarms is common in ICUs, its use has not been well studied for predictive alarms. The paper notes this may be due to the inability of classic metrics to correctly assess the impact of snoozing. The u-metrics solution correctly assesses the impact of snoozing because it will neither reward nor penalize the system for suppressing alerts that were correct but would create distraction or harm if fired. The paper also describes a method to identify optimal snooze times. Properly applied, snoozing may reduce alert fatigue and increase the likelihood that clinicians will respond to true alarms.

More information: Jonathan A. Handler et al, Novel Techniques to Assess Predictive Systems and Reduce Their Alarm Burden, *IEEE Journal of Biomedical and Health Informatics* (2022). DOI: 10.1109/JBHI.2022.3189312

Provided by OSF Healthcare