Title
Correlation and crowding measures: the fundamental lesson behind complex statistics

Permalink
https://escholarship.org/uc/item/7qv067v7

Author
Weber, Ellen J

Peer reviewed

Correlation and crowding measures: The fundamental lesson behind complex statistics.

As someone who has taken basic statistics three times, I might be the least likely person to write this commentary highlighting the findings by Boyle and colleagues in their paper "Comparison of the International Crowding Measure in Emergency Departments (ICMED) and the National Emergency Department Overcrowding Score (NEDOCS) to measure emergency department crowding: pilot study. [1] But then, perhaps having a statistical novice such as myself providing a synopsis of what these authors have done will demonstrate that, despite a somewhat heady methods section, the underlying concepts are as simple to grasp as they are critically important.

 The authors set out to compare the ability of the 7-point International Crowding Measure in Emergency Departments (sICMED) with the well-validated NEDOCS score, which originated in the US, to reflect the sense of crowding and danger by senior physicians in England. To do this, one of the investigators collected data at four different emergency departments for each of the scores every hour, and asked the senior physician on duty at the time to rate their sense of danger and crowding at the same time. The investigators then compared the ability of the two scores to predict the senior physicians' impression. This is the design of many studies, including one by the authors, in determining how well crowding scores reflect the reality on the ground. And when Boyle et al performed this analysis in the current

study, they found that both the sICMED and the NEDOCS had statistically significant

associations with the outcome (perception of danger and crowding)

Simple enough. In fact, too simple.. There was a problem: observations set closely in

time (like every hour) are likely to be highly correlated. If your department has 30

treatment spaces, for example, and at 2PM all 30 are full with 10 patients in the

waiting room, its unlikely that at 3PM occupancy will be much less than 40, or that it

was much less at 1PM. We all know it takes hours (sometimes days) to decant the

department to a reasonable size.

A basic assumption of simple statistical tests for hypothesis testing is that the

observations in your data set are independent – that the observation at one time has

no relationship to another observation. If the observations are not independent,

then you have a lot of data saying (nearly) the same thing and using statistical

methods that assume such independence   will falsely elevate the precision of your

estimates, potentially leading to erroneous conclusions of statistical significance. In

essence, if you count each observation as independent, you have falsely elevated

your N, and your power to detect a small difference–what Boyle and colleagues call

"the naïve result."

In their study, Boyle and colleagues conducted 82 hourly observations conducted on

10 separate days among four EDs. To begin with they worked with the 82

observations, and applied a series of increasingly complex regression models each

designed to make assumptions that were more consistent with what is likely to be

true about the underlying process generating the data. With each successive model,

they found smaller and smaller effect sizes, so that by the fourth and most complex model (a time series which accounted for the correlation structure and the time of day of the observations), no statistical association was found between SICMEDS and danger or crowding, while the precision of the association of for NEDOCS with the outcomes was reduced (but still significant). As a final model, which perhaps is easiest for all of us to understand, they looked at the data as only 10 observations (as the hourly observations within each were correlated) and adjusted for clustering by site; in this analysis, there was no statistically significant association between either score and the perception of crowding, although an association with perception of danger remained.

It is not necessary for most readers to fully understand every step of the analysis in this article, but it is important to recognize that when one correctly accounts for correlation of observations, statistical associations are weakened. This should make us review carefully previously published crowding literature for whether correlation between measurements made closely in time was accounted for. More importantly, all future studies on crowding need to take this very important (and basic) statistical concept into consideration.

1References

Boyle A, Abel G, Raut P et al. Comparison of the International Crowding Measure in Emergency Departments (ICMED) and the National Emergency Department Overcrowding Score (NEDOCS) to measure emergency department crowding: pilot study.2015 Emergency Medicine Journal epub ahead of print