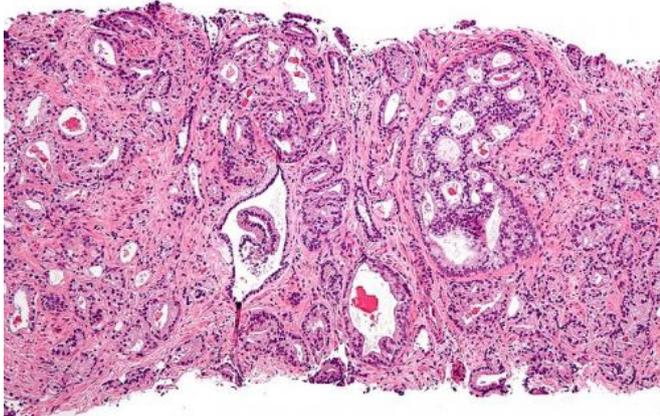# A mathematical crystal ball gazes into future of prostate cancer treatment

5 August 2017

Micrograph showing prostatic acinar adenocarcinoma (the most common form of prostate cancer) Credit: Wikipedia

The chemotherapy docetaxel is widely accepted as a standard therapy for metastatic castration-resistant prostate cancer. But 10-20 percent of patients will have adverse side effects that force discontinuation of treatment. These patients may have been better off with another treatment or alternative dosing of docetaxel in the first place, but who's to know before trying the drug which patients will go on to experience debilitating side effects? A crowdsourced competition asked this as an open question. Today in the *Journal of Clinical Oncology Clinical Cancer Informatics*, competition organizers and participating teams report their findings: Using open data from four previously conducted clinical trials, teams of international researchers designed mathematical models predicting the likelihood that a patient will discontinue docetaxel treatment due to adverse events. These results represent the first comprehensive effort to make such predictions based on patient clinical characteristics.

Specifically, the challenge was to connect any of

129 baseline clinical measurements to the chance of docetaxel discontinuation. In all, 34 international teams submitted 61 models. Seven of these teams submitted models with similarly high predictive ability and so technically "won" the challenge. The five clinical factors that were most predictive were measures of hemoglobin, alkaline phosphatase, aspartate aminotransferase, prostate specific antigen, and ECOG performance status. The seven successful models all integrated these five factors into various computational frameworks.

Interestingly, after the competition officially ended, these top seven teams decided to collaborate outside the framework of the competition, resulting in refinements that led to a combined model that was more predictive than any of the submissions alone.

"The seven groups from around the world—Finland, Germany, Canada, Israel and the U.S.—had never formally met before the challenge. It's a really exciting example of the power of scientific collaboration," says James Costello, PhD, senior author of the paper, investigator at the University of Colorado Cancer Center, assistant professor in the Department of Pharmacology at the CU School of Medicine, and director of Computational and Systems Biology Challenges within the Sage Bionetworks/DREAM organization.

The combined model stratified patients into groups with low and high risk of discontinuing docetaxel due to adverse events, with the high group having more than double the likelihood of discontinuation as the low group.

"Not only could a model like this help identify patients who might benefit more from a different treatment, it also has the potential to immediately impact future clinical trials by improving patient selection through the use of novel patient selection designs. In doing so, the number of patients needed for clinical trials could be reduced, making

more efficient use of available resources," says Devin Koestler, PhD, assistant professor of Biostatistics at the University Kansas Medical Center, and one of the first authors of the paper.

The challenge was built to promote and capitalize on the potential of an open question paired with open data.

"The field is definitely moving toward a much more open sharing model of clinical trial data. This project is a great example of how you can gain new knowledge from existing data and how making clinical data open and freely accessible can maximize the use of these valuable data for the benefit of patients," says Laura Elo, PhD, one of the senior authors of the paper, adjunct professor in Biomathematics and research director in Computational Biomedicine and Bioinformatics at Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Finland.

The project was overseen as a collaborative effort between 16 institutions, led by academic research institutions including CU Cancer Center, open-data initiatives including Project Data Sphere, Sage Bionetworks, and the DREAM Challenges, and industry and research partners including Sanofi, AstraZeneca, and the Prostate Cancer Foundation.

Because only 10-20 percent of patients discontinue treatment due to adverse events, no single trial has enrolled enough patients to predict with statistical significance who would discontinue docetaxel - commonly, these trials tested the effectiveness of treatments in the population that was able to finish the regimen and were not designed to answer this secondary question of who would be unable to finish. Had clinical trial results remained firewalled by the academic or industry sponsors, this secondary question would have remained unanswered; however, the decision to open these clinical trial data allowed the current researchers to combine the numbers from four previous trials, pooling over 2,000 patients - enough to start identifying statistically significant patterns.

"The number of clinical trials in Project Data Sphere continue to grow. At the time of this study, there were about 10,000 patients in their database. Now

there are over 70,000, meaning that we will be able to explore future questions with even greater accuracy and ask questions that have been unaddressable due to restricted data access," Costello says.

The project takes place in the context of a debate between research factions, one of which holds that studies should generate new data designed to specifically explore study questions, and another faction that sees value in mining previously generated data for new insights. While antagonizing terms, such as "research parasites," have been used to describe such data scientists, this study bridges the gap between these factions to develop a necessary research symbiosis. The clinicians who helped generate the original data from clinical trials worked with the researchers to develop new tools and draw novel conclusions from these data.

"Ultimately what we'd like to do is have a much more dynamic interaction with the clinical trial design. As patients are coming in, we could use these statistical models to help match the right person with the right trial," Costello says.

Chalk one up for research symbiosis: International, multi-institution research collaboration to mine previously generated, open data now offers a tool that can better leverage patient clinical characteristics and ease the process of clinical trials.

Provided by CU Anschutz Medical Campus