

Scientists don't have to travel alone; solutions can come from the crowd

Science is sometimes perceived as a lonely affair where a lone genius strives to break the code that will unlock hidden knowledge deep inside the data. Without question the contributions of great scientists are of value, but alternatives are thriving as well.

In our article recently published in the *Lancet Oncology*, "Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: development of a prognostic model through a crowdsourced challenge with open clinical trial data", we present a novel solution to the difficult and highly clinically relevant problem of determining prognosis of patients with metastatic castration-resistant prostate cancer treated with standard chemotoxic therapy. A gifted statistician with extensive experience in statistical modeling in oncology developed the prior state-of-the-art method.

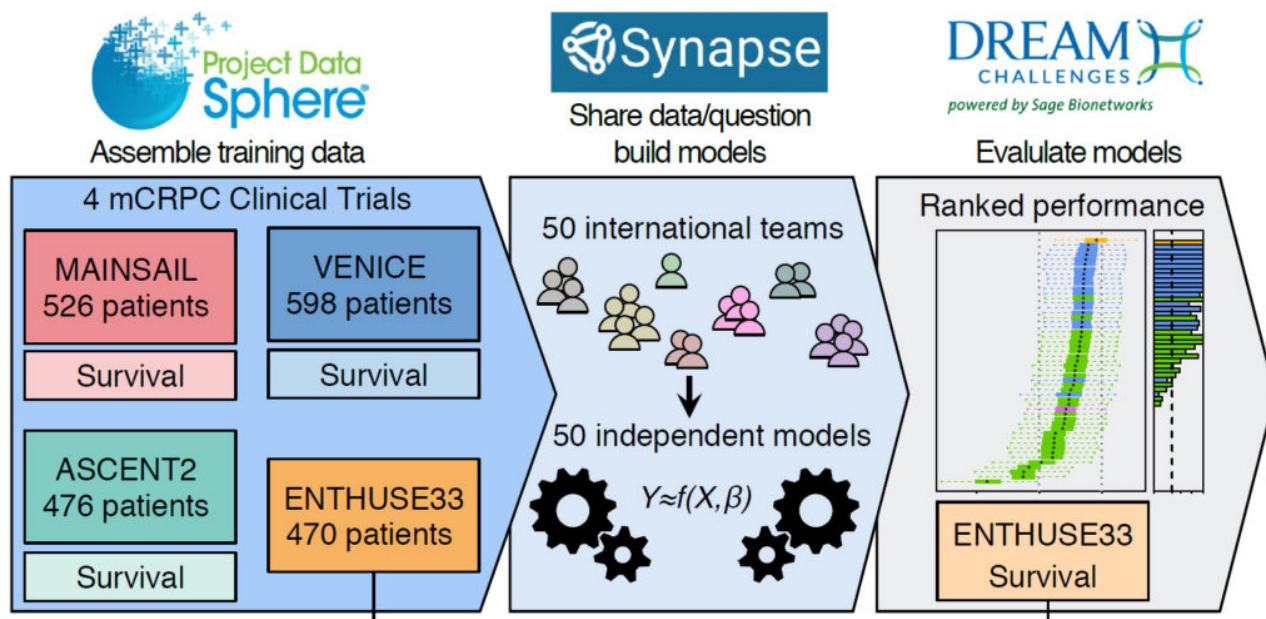


Fig. 1. Overview of the Challenge to build prognostic models for patients with metastatic, castration-resistant prostate cancer treated with standard chemotherapy.

In our Challenge, we recruited a global crowd of data scientists, biomedical researchers, engineers, clinicians, and biostatisticians to develop the best performing prognostic model. The top-performing model was developed by a group from Finland with no prior experience in predictive modelling of prostate cancer patient survival. How could this be?

First, a Challenge was issued (by the DREAM Challenges and Sage Bionetworks) and a prize was announced via a cloud-based, data sharing and analysis platform (Synapse). Of note, the prize was far lower than the cost of a typical grant. In the global age of the internet, making potential investigators aware of the “Challenge” and “the prize” is much easier now than in the past.

Second, an objective methodology for determining model performance was determined prior to the Challenge. This measure of success was statistically defined and benchmarked against the above-mentioned long-standing and well-respected reference model. This ensured a fair comparison of all the submitted entries, and made sure that the Challenge would result in clinically interesting improvements for modeling patient survival.

Third, a data set was made available and freely shared to all interested in the Challenge. This exceptionally large, real-world prostate cancer clinical trial data was the product of a fruitful collaboration between some of the world’s largest pharmaceutical firms and academic institutions.

The Challenge was done in two phases: firstly, the teams were able to test their methods against limited test data set; secondly teams were scored and ranked on a separate, independent evaluation data set.

The results were both surprising and gratifying. A total of 50 international teams consisting of hundreds of individuals submitted models, and the sophistication of the responses was immediately evident. More than half of the submitted models bested the prior reference model, indicating substantial expertise (despite many of the groups having no prior recognition or expertise in cancer research). The top-performing model was a clear winner, and statistically significantly better than the standard reference model as well as the other competing models.

What did we learn? First, that crowdsourcing can bring out experts who were eager to tackle such challenging research questions when given the chance. Second, the crowd-sourced solution was superb, better than the prior reference standard. Third, the cost of the actual prize was comparatively low relative to the standard system of dispensing grants to individuals (which can promise only uncertain outcomes).

What was key to the success of this endeavor was the availability of clean, robust and freely available clinical trial data. This was absolutely vital to the success of this Challenge. The data was provided as part of a larger data sharing initiative -Project Data Sphere, which makes detailed clinical trial data available in the public domain.

Through open data-sharing and crowdsourcing, we enabled the wisdom of many to come together to contribute to the prediction of prognosis in a common cancer associated with worldwide suffering. Making progress in science is not just up to talented lone wolves - it is best when shared among the body of talented international researchers!

Oliver Sartor¹, Teemu D Laajala^{2,3}, Justin Guinney⁴, Tao Wang^{5,6}, Martin J. Murphy⁷, Tero Aittokallio^{2,3}, Fang Liz Zhou⁸, James C Costello⁹

¹*Tulane Cancer Center, Tulane University, New Orleans, LA, USA*

²*Department of Mathematics and Statistics, University of Turku, Finland*

³*Institute for Molecular Medicine Finland, FIMM, University of Helsinki, Finland*

⁴*Sage Bionetworks, Seattle, WA, USA*

⁵*Quantitative Biomedical Research Center, Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA*

⁶*Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX, USA*

⁷*Project Data Sphere®, LLC, Cary, NC, USA*

⁸*Sanofi, Bridgewater, NJ, USA*

⁹*Department of Pharmacology, University of Colorado, Anschutz Medical Campus, Aurora, CO, USA*

Publication

[Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: development of a prognostic model through a crowdsourced challenge with open clinical trial data.](#)

Guinney J, Wang T, Laajala TD, Winner KK, Bare JC, Neto EC, Khan SA, Peddinti G, Airola A, Pahikkala T, Mirtti T, Yu T, Bot BM, Shen L, Abdallah K, Norman T, Friend S, Stolovitzky G1 Soule H, Sweeney CJ, Ryan CJ, Scher HI, Sartor O, Xie Y, Aittokallio T, Zhou FL, Costello JC; Prostate Cancer Challenge DREAM Community.

Lancet Oncol. 2017 Jan