

Data Sharing, Research Parasites, and the Korean Alphabet

Faculty Presenter

Geoffrey Kim, MD, Director, Division of Oncology Products 1, Office of Hematology and Oncology Products, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA

Scholar Summary

*Authored by **Sebastiano Battaglia, PhD**, Roswell Park Cancer Institute Center for Immunotherapy, Buffalo, NY, USA*

An interesting editorial from Dan L. Longo and Jeffrey M. Drazen was published on January 2016 on the New England Journal of Medicine (NEJM). The authors categorized as “research parasites” those investigators and researchers who utilize publicly available datasets to run their own research without prior direct input in the data collection.

Geoffrey Kim offered an interesting perspective on this editorial by leveraging the origins of the Korean alphabet. The 3rd king of Joseon Dynasty, Sejong the Great, strongly supported the development and spreading of the Hangul Alphabet, or Korean Alphabet. In ancient times reading and writing was a luxury, and Sejong the Great developed a very clever additive system to build phonetic elements on simplified blocks of the Chinese alphabet. The simplicity of this approach allowed everyone, including farmers living in rural areas, to learn how to read and write. However, as the NEJM authors complained about “research parasites”, the Korean upper classes complained about people in lower classes having access to the Hangul. As Sejong the Great developed a new alphabet accessible to everyone, Geoffrey Kim at the FDA, under the direction of Martin J. Murphy, developed the Project Data Sphere (<https://www.projectdatasphere.org/projectdatasphere/html/about>) with the goal of allowing investigators to share, access and analyze clinical trials data.

Upon being approved for access, investigators can leverage an integrative platform to visualize and analyze trials data with SAS. As of December 2016, there are 69 trials available, containing patient-level data from phase III clinical trials across numerous tumor types. Furthermore, the Project Data Sphere Consortium is currently working on including data from immunotherapy trials.

Overall, this is a milestone project that, like the Cancer Genome Atlas (TCGA), allows investigators around the world to test and develop new hypotheses, validate data, and establish collaborations with the overarching goal of increasing scientific knowledge and improving cancer treatment worldwide.