Cancer research is awash with databases that capture widely different aspects of the disease, from tumor samples and genomic sequencing to clinical study results, socioeconomic, and billing. This information typically gets warehoused in disconnected data sets, investigated by different types of researchers, and published in journals specific to their focus of research. Michaela Dinan, PhD, Associate Professor of Epidemiology (Epidemic Diseases) and Co-Leader of the Cancer Prevention and Control Research Program at Yale Cancer Center, notes that “linking disparate data sets is a lost opportunity. Her research demonstrates that when disparate data sets are pushed into conversation with each other, they can disclose new insights about cancer, cancer care, and the healthcare system.”

“If you think of novel ways to use data that have been around a long time,” said Dr. Dinan, “you can make real contributions to the field.”

Her most recent contribution was published in *JAMA Network Open* in October 2021. The paper describes a pilot study that investigated how breast cancer screening in October 2021. The paper describes a pilot study that investigated how breast cancer screening in resource-poor neighborhoods with less access to care may be important as well.”

The combined data revealed that socioeconomic status and access to screening remained associated with mortality among patients with breast cancer. “That’s probably our number one finding,” said Dr. Dinan. “Our research suggests that living in resource-poor neighborhoods with less access to care may be important as well.”

To link and cross-analyze databases might seem obvious in retrospect, but it had been done before. To collect and interpret the merged data took almost a decade. “There were lots of roadblocks,” she said. “One of the main challenges in obtaining funding was the concern that ‘It isn’t feasible.’ But now we can say it’s possible because we’ve done it.”

Dr. Dinan is now proposing the first-ever linkage between the SEER-Medicare databases and the SEER Virtual Tissue Repository (VTR), which is a prospective, forward-facing version of the work done with the SEER-RTR. Dr. Dinan wants to mine the databases to answer two questions about the use of immunotherapy to treat renal cell carcinoma (RCC). About 20 percent of RCC patients have a “durable response” to these therapies, meaning a potential cure, but no one can predict who these patients will be. Instead, between one to three percent of RCC patients have severe toxic reactions to immunotherapies. Again, no one knows beforehand who these patients will be. This is where Dr. Dinan’s methodology shows its value.

Dr. Dinan will use the SEER-Medicare data to identify everybody who received immunotherapy for RCC and identify two subsets of patients, one that shows evidence of a durable response and another that shows evidence of a severe reaction. This will allow her to study what’s different about these patients. She’ll then use the SEER-VTR to identify two cohorts of patients, one that shows evidence of a durable response and another that shows evidence of a severe reaction and then examine a variety of socioeconomic factors that may drive these rare events, whether a durable response or a severe reaction. This will allow her to study what’s different about these patients.

Michaela Dinan, PhD
Yale Cancer Center
19yalecancercenter.org |