Cancer research is awash with databases that capture widely different aspects of the disease, from tumor samples and genomic sequencing to clinical study results, sociodemographics, 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 (Epidemiologic Disease) and Co-Leader of the Cancer Prevention and Control Research Program at Yale Cancer Center, sees that this not only is frustrating but that research demonstrates that when disparate data sets are pulled into conversation with each other, they can excite 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 impacts clinical, genomic, and sociodemographic factors associated with newly diagnosed breast cancer. Dr. Dinan and colleagues did this by combining and cross-analyzing associated with newly diagnosed breast cancer. All patients have their identity protected by extensive checks and balances to ensure that no individual patient can ever be identified from the research. The novelty of the project is that Dr. Dinan and her colleagues combined this SEER-Medicare data with physical tumor specimens from the SEER-Residual Tissue Repositories (RTRs) and then conducted gene expression analysis.

“This was the first study to link physical tumor samples from these patients in the SEER database to Medicare claims data, and to create one novel data set,” said Dr. Dinan. “If you only look at one data set, you’re not getting the whole picture.”

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 Dr. Dinan understand part of the reason it had not been done before. To collect and interpret the target 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 database 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. Second, between one to three percent of RCC patients have severe toxic reactions to immunotherapies. Again, no one knows beforehand who these patients will be.

“From 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 cohorts of patients, one that shows evidence of a durable response and another that shows evidence of a severe adverse event and another that shows evidence of a severe adverse event.

“Next, we’ll cherry pick these people,” said Dr. Dinan. “This is where Dr. Dinan’s methodology shows its value.”

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 database 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. Second, between one to three percent of RCC patients have severe toxic reactions to immunotherapies. Again, no one knows beforehand who these patients will be.

“Instead of waiting to see what happens in a clinical trial, we’re going to find the outcomes of interest first, and then go back and pull those patients to study what’s different about them.”

“Instead of waiting to see what happens in a clinical trial, we’re going to find the outcomes of interest first, and then go back and pull those patients to study what’s different about them.”

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 database 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. Second, between one to three percent of RCC patients have severe toxic reactions to immunotherapies. Again, no one knows beforehand who these patients will be.

“Instead of waiting to see what happens in a clinical trial, we’re going to find the outcomes of interest first, and then go back and pull those patients to study what’s different about them.”

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 database 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. Second, between one to three percent of RCC patients have severe toxic reactions to immunotherapies. Again, no one knows beforehand who these patients will be.

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 database 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. Second, between one to three percent of RCC patients have severe toxic reactions to immunotherapies. Again, no one knows beforehand who these patients will be.

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 database 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. Second, between one to three percent of RCC patients have severe toxic reactions to immunotherapies. Again, no one knows beforehand who these patients will be.