

researchblog.duke.edu

Diabetes - and Privacy - Meet 'Big Data' - Research Blog

Maya Iskandarani

4–5 minutes

“Click [here](#) to consent forever.”

If consent to participate in medical research were that simple, [Joanna Radin](#) of Yale University would have to find a new focus for her research, and I would never have found the [Trent Center for Bioethics, Humanities & History of Medicine](#).

Luckily for us both, this is not the case. Medical consent is a very complex issue that can, as Radin’s research attests, traverse generations.



Joanna Radin’s research focuses on the intersection of medical

history, anthropology and ethics at Yale University. Source: Yale School of Medicine

Radin is an Associate Professor of Medical History at Yale, the perfect fit for the [Humanities in Medicine Lecture Series](#) taking place this month at the Trent Center. Her research nails the narrow intersection of medical history, anthropology, bioethics and data analytics. In fact, Radin's appeal is so broad that her visit to Duke was sponsored by no less than six Duke departments, including the Departments of Computer Science, History, Electrical and Computer Engineering, Cultural Anthropology and Statistical Science.

Radin's lecture honed in on a well-known case in the realm of bioethics and medical history: the [Pima Native American](#) tribe in Arizona, which is known for unusually high rates of diabetes and obesity. The Pima were the first Native American tribe to be granted a reservation in Arizona—30,000 acres—at the beginning of the California Gold Rush. In 1963, following nearly half a century of mass famine among the Pima, the National Institute of Health (NIH) conducted a survey for rheumatoid arthritis in the Pima tribe, instead discovering a frighteningly high frequency of diabetes.

In 1965, the NIH initiated a long-term observational study of the Pima that continued for about 40 years, though it was meant to last no more than 10. The goal of the study was to learn about diabetes in the “natural laboratory” of sorts that the Pima reservation unwittingly provided. The data collected in this study came to be known as the Pima Indian Diabetes Data set (PIDD).

Machine learning enters the story around 1987, when David Aha and colleagues at the University of California, Irvine (UCI) created the [UCI Machine Learning Repository](#), an archive containing thousands of data sets, databases and data generators. The repository is still active today, virtually a gold mine for researchers in machine learning to test their algorithms. The PIDD is one of the oldest data sets on file in the UCI archive, “a standard for testing data mining algorithms for accuracy in predicting diabetes,”

according to Radin.



A Pima farmer in Pima, Arizona, circa 1900. Source: Wikimedia Commons

Generations' worth of data on the Pima tribe have been publicly accessible in the UCI archive for over two decades, creating ethical controversy around the accessibility of information as personal as blood pressure, body mass index (BMI) and number of pregnancies of Pima Native Americans. Though the PIDD can help refine machine learning algorithms that could accurately predict—and prevent—diabetes, the privacy issues provoked by the publicness of the data are impossible to ignore.

This is where “eternal” medical consent enters the equation: no researcher can realistically inform a study participant of what their medical data will be used for 40 years in the future.

These are the interdisciplinary questions that Radin brought forth in her lecture, weaving together seemingly opposite fields of study in an engaging, thought-provoking presentation. No one who left that room will look at the Apple Terms & Conditions the same way

again.

Post by Maya Iskandarani
iskandarani_maya_100hed