DTBC Speech

San Francisco 2019

First off, I want to thank the members of the DTBC Executive Board—Josh Freilich, Steve Chermak, Carla Lewandowski, Sue Ming Yang, Nancy Morris, Pete Simi, Will Parkin, Leevia Dillon.

--What a nice surprise!

I also, want to thank my son Andrew and his lovely wife Talia for joining the event today—they flew up from San Diego.

--I thought in the brief time I have I would talk a little bit about how I ended up doing research on terrorism and political violence.

--But I also wanted to be a little bit controversial—I also want to talk a little bit about large archival databases on terrorism like the Global Terrorism Database, the PIRUS data, the ECDB data and the American Terrorism Study.

I promise to keep it short--the organizers asked me to talk for about 15 minutes.

--I started out as an undergraduate at Indiana University studying history—only discovered criminology during the end of my junior year.

--I think a background in history has had a lasting influence on my career in Criminology—have sometimes thought secretly that it has put me at odds with mainstream American criminology

--the dominant method for studying crime has probably been the survey.

--but like many historians, I have always been more trustful of data based on the records people leave behind than what they say about themselves.

--I have published research based on surveys but I have always been a bit nervous about survey data—surveys seem so ephemeral, dependent on the characteristics of the interviewer, the social connection between the interviewer and the interviewed, the time of day, the time of year, the order of the questions, and the interviewee’s presentation of self—and so on.

--During graduate school I came to believe that the most defensible route for criminology was for researchers to become what could be called social “archeologists”—that is, to look for archival data left behind by others that might shed light on key questions about deviance and crime.

--As in archeology, we would approach these data realizing that they are incomplete but expecting that based on interpreting the bits and pieces of this admittedly imperfect data, we could arrive at an assessment of the underlying reality. Like archeologists, reconstruct an artifact by looking at the pot shards left behind.

-- For the past three decades, this archeological approach has led me to spend most of my academic career looking for, digitizing, and analyzing archival data sets on crime and violence.

--In the early part of my career I put together large archival databases on robbery, burglary and rape.

--During the middle part of my career I did a lot of work putting together large cross-national databases on homicide.

And the discovery of a database on terrorist attacks in 2001 led me to work—for the past twenty years—on archival databases on terrorism.

The most extensive of these archival databases on terrorism—the Global Terrorism Database—was a major factor in the launch of the National Consortium for the Study of Terrorism and Responses to Terrorism—the START Center—which I helped create and then directed for 14 years.

But starting each new project with archival data frequently runs you into conflict with the dominant paradigm of academic criminology: namely, that the way we are to advance our field is through theoretical development and deductive reasoning.

--And here I want to admit some things that I have never publicly admitted—I hope there are no journal editors in the audience!

--When you start a project with an archival dataset and ask questions of this data set, it is difficult—or maybe even impossible—to do deductive, theoretically based science.

--The problem is that it is very hard to publish in any of the top ranked journals in criminology or the other social sciences if you do not start the article with a set of theoretical expectations.

You all know the drill—introduction, theoretical conceptualization, literature review and hypotheses.

-- We have been taught that the way criminology should work is that we start with a theory, develop hypotheses, collect data, and then test whether the theoretical model is supported or falsified.

--But this model is complicated when you start the research with an already collected data base.

So, how have I been publishing articles in criminology for the past 30 years—frequently by doing the analysis first, then figuring out what the data say, and then selecting the theory that seems to fit the data the best.

--so I have been writing a lot of papers in the opposite way that is recommended by the mainstream criminology paradigm: I write the methods and results first and the theoretical conceptualization last!

--okay—I am finally coming out of the closet publicly on this because I am getting increasingly convinced that these inductive methods are not all bad.

--that criminology-and in particular, research on terrorism and bias crime-- is making a mistake if it emphasizes deductive, theory-driven research too much. Here are three reasons why.

1. First, many of the most important policy questions we face in the study of terrorism and bias crime are descriptive and have little to do with theory.

--When we created the GTD in the early 2000s, no one on the planet could tell how many terrorist attacks were occurring worldwide? How many individuals were being killed by terrorists? What groups were responsible? What types of weapons were being used? Were the total number of attacks going up or down? What regions of the world, countries and cities had the most frequent attacks? And so forth…

--note that answers to these questions do not require anomie, social control, differential association or any other theory.

--they are simply descriptive questions.

--I still remember when we were finally able to produce our first version of the GTD, which covered the years 1970 to 1997—seeing the trend lines made me feel like the rocket scientists from NASA must have felt when they watched astronauts walking on the moon.

--for the first time in human history a group of researchers had the ability to try and recode detailed information on every single terrorist attack happening around the world.

--I still get a bit of this feeling today when Erin Miller produces these great longitudinal charts showing the latest wave of GTD data

2**. Second, sometimes correlation is good enough.**

--I know—for many people in this room, this may be a controversial statement.

--Lots of times knowing that something works without understanding why it works is still extremely important.

--With the growing importance of big data we are seeing an increasing emphasis on correlations between things which occur together in a way not expected on the basis of chance.

--Thirty years ago the statistician George Box said that “all models are wrong, but some are useful.”

--for example, if we know that a particular kind of drug is effective in stopping heart attacks, we do not necessarily need to know how it works.

--knowing that the vast majority of terrorist attacks are domestic attacks, or that convicted terrorists who have spent time in prison are more likely than others to be violent can be very useful—even if we cannot yet say why.

--but even this statement is not radical enough because it implies that correlation is just a stepping stone to causation—sometimes it may be the point.

--for example, Google has been able to provide those fantastic tools for translating languages without knowing a thing about how languages actually operate—but only by analyzing millions of translation cases and using algorithms that make these translations progressively more accurate.

--the same type of technology allows Tesla to develop increasingly sophisticated self-driving cars without any theoretical models of how humans operate automobiles.

--in fact, Peter Norvig, Google’s research director, recently offered an update to George Box’s maxim: “All models are wrong, but increasingly, you can succeed without them.”

--perhaps the best practical example of this is the gene sequencing work of Craig Venter and his colleagues.

--enabled by high speed sequencers and supercomputers, Venter went from sequencing individual organisms to sequencing entire eco systems.

--in the process he is discovering thousands of previously unknown species of bacteria and other life forms.

--Venter can tell you almost nothing about the species found—what they look like, how they live or what is their morphology

--He just has lots of data.

--but if these data are correlated with other gene sequences of species he can make some educated guesses about the animals referenced—can they convert sunlight into energy, are they descended from a common ancestor, etc.

--with my colleagues, Scott Decker, David Pyrooz, Michael Becker and Patrick James, we recently did some work comparing how individuals become gang members or political extremists.

--We have produced some correlations that have interesting policy implications even though the work is atheoretical.

--For example, only about 6 percent of gang members go on to become political extremists.

--Both gang members and extremists are likely to come from broken families

--while gang members stress material rewards as the reason they join gangs, political extremists stress moral rewards

--correlations may not tell us why something is happening, but they may alert us that something is happening—and in many situations this is enough.

**Finally, point number 3. Doing research on populations rather than samples has important implications for the utility of theory.**

--theory is often of greatest use in terms of extrapolating from samples to populations.

--in the past, criminology has most often dealt with samples—the National Crime Victim Survey, the proportion of countries of the world that collect and report homicide data.

--one of the things that is revolutionary about the GTD is that it includes all terrorist attacks from everywhere on the planet—the whole population of terrorist attacks.

--having access to whole populations is cropping up more frequently in many areas of social science study—the universe of Twitter users; the universe of people using Google Search; open source data bases on all school shootings or all shootings of civilians by police officers.

--so as we develop bigger and more inclusive data bases that approximate populations rather than samples, our understanding of the underlying reality can hopefully evolve.

--we see this reasoning playing out especially with regard to bioinformatics.

--The process of model building in biology is increasingly driven by the massive amount of data produced and is correspondingly, less dependent on theories and hypotheses.

--One pioneer in this field put it this way: “the goal is to discover things we neither knew nor expected, and to see the relationships and connections among the elements, whether previously suspected or not.”

--the best inductive algorithms can evolve—for example, the team that is collecting the GTD has used machine learning to refine their methods of collecting data—their data collection system evolves.

--no one on the GTD would presume to argue that the data are perfect—but in some ways, learning is always incomplete.

--permanent learning, never completed, produces an imperfect but useful knowledge.

Conclusions

--Now I do not want to push these arguments too far.

--Richard Feynman, an American theoretical physicist, wrote on his blackboard shortly before his death, “What I cannot create, I cannot understand.”

--For Feynman, truly understanding something meant being able to understand each step of the process.

--correlation alone cannot do this—we need to follow up with models and experiments—for example, distinguish between meaningful and spurious results.

--more data do not necessarily generate more knowledge.

--Data by themselves are meaningless.

--And remember that data collection itself is often based on prior theories and knowledge

--So, I am not going to make the argument that theory is dead or that inductive methods will completely replace—or should completely replace--deductive hypothesis testing.

--But I would say that the increasing development of extensive worldwide data bases on topics such as terrorism are likely to become increasingly important and can be an extremely useful source of information for criminology and public policy more generally.

Thank you!