

Posted on Mon, Nov. 27, 2006

Phila. getting software to predict who might kill

By Michael Matza
Inquirer Staff Writer

University of Pennsylvania criminologist Richard Berk, a trained statistician, never met a data set he didn't like.

Now, using fresh data from the Philadelphia probation department, Berk and three colleagues have built an innovative model for predicting which troublemakers already in the system are most likely to kill or attempt a killing.

With the homicide rate in Philadelphia outpacing last year's by at least 7 percent, a computer model for "forecasting murder" is in the works, Berk said, to be delivered to the probation department in the new year, with clinical trials of the new tool to begin in the spring.

Initial research suggests the software-based system can make it 40 times more likely for caseworkers to accurately predict future lethality than they can using current practices.

The project is funded with a private grant and the software is in the public domain, so the product will be delivered to the city free.

"This will help stratify our caseload and target our resources to the most dangerous people," probation department director of research Ellen Kurtz said. "I don't care as much about [targeting] the shoplifter. I care a lot about the murderer, obviously."

Sociologists have produced hundreds of studies using discrete pieces of information about offenders to try to come up with the means to identify probationers most likely to commit future felonies.

But homicide, because it is a relatively rare event, has been very hard to predict. Of all probationers in Philadelphia, only about one in 100 will commit homicide. But for obvious reasons it is crucial to find that needle in the haystack, Berk said.

Though it is well known that the probability of becoming a killer falls off with age, especially between 18 and 30, Berk's innovation is that he uses the relatively new statistical technique called "step-function" analysis to show how fast and when the dropoffs occur.

"In reality the risk doesn't decline in a smooth, straight line" but falls precipitously at certain points for certain reasons, he said.

The tool works by plugging 30 to 40 variables into a computerized checklist, which in turn produces a score associated with future lethality.

"You can imagine the indicators that might incline someone toward violence: youth; having committed a serious crime at an early age; being a man rather than a woman, and so on. Each, by itself, probably isn't going to make a person pull the trigger. But put them all together and you've got a perfect storm of forces for violence," Berk said.

Asked which, if any, indicators stood out as reliable predictors of homicide, Berk pointed to one in particular: youthful exposure to violence.

Using probation department cases entered into the system between 2002 and 2004, Berk and his colleagues performed a two-year follow-up study - enough time, they theorized, for a person to reoffend if he was going to. They tracked each individual, with particular attention to the people who went on to kill. That created the model. What remains at this stage is to find a way to marry the software to the probation department's information technology system.

When caseworkers begin applying the model next year they will input data about their individual cases - what Berk calls "dropping 'Joe' down the model" - to come up with scores that will allow the caseworkers to assign the most intense supervision to the riskiest cases.

Even a crime as serious as aggravated assault - pistol whipping, for example - "might not mean that much" if the first-time offender is 30, but it is an "alarming indicator" in a first-time offender who is 18, Berk said.

The model was built using adult probation data stripped of personal identifying information for confidentiality. Berk thinks it could be an even more powerful diagnostic tool if he could have access to similarly anonymous juvenile records.

The central public policy question in all of this is a resource allocation problem. With not enough resources to go around, overloaded case workers have to cull their cases to find the ones in most urgent need of attention - the so-called true positives, as epidemiologists say.

But before that can begin in earnest, the public has to decide how many false positives it can afford in order to head off future killers, and how many false negatives (seemingly nonviolent people who nevertheless go on to kill) it is willing to risk to narrow the false positive pool.

"If we have 100 probationers I can accurately find the one murderer who statistically will be in that group if I devote resources to all 100 as if they are murderers. The problem is that for that one murderer who is a 'true positive,' I have 99 false positives. We all would agree that's not a good use of resources.

"Now suppose I can identify the 10 at highest risk. For that one true positive I now have nine false positives," Berk said, "and that may be something we choose to live with."

Contact staff writer Michael Matza at 215-854-2541 or mmatza@phillynews.com.