

Old Saws and Modern Instances: On not making inferences based on “group data”

Vernon L. Quinsey¹
Queen's University, Kingston

I always wondered what, “Happy as a clam,” meant, especially because my empathic and theory of mind skills are insufficient to penetrate clamshells. When I discovered that the complete phrase is, “Happy as a clam at high-tide,” I got it. I’ve had a similar experience about the admonition not to make an inference or prediction about an individual from “group data.” It sounds sort of wise until you start thinking about what it could actually mean. I’ve puzzled about it off and on for years without ever really getting it – until now!

I’ll illustrate the problem with a parable provided to me by my friend, Grant Harris, a long-time, part-time resident of the Oak Ridge Division of the Mental Health Centre, Penetanguishene. He wrote asking for bail money on his Blackberry after an unfortunate incident with a car salesman.

Being an aspiring sporty guy, Grant was shopping for a Subaru Outback. The salesman told him about the repair record of these Outbacks, based on many cars of the same model. Grant said, “Not so fast; I only want to buy one particular car. How can I know whether this individual car is not a lemon?” The salesman replied that the odds were very high that this individual car was good. “But,” said Grant, looking smug, “Those odds are based on group data -- group data are useless to me; I need to know about this individual vehicle, and all those group data say is that it’s not a certainty that this car is a lemon. Otherwise, I know essentially nothing about this car. It’s obvious that cars, even of the same model and the risk of lemons, are not homogeneous.” The salesman smiled warily and replied that the Outback was suitable for all lifestyles.

Grant couldn’t decide whether to give the salesman a cheque for \$30K or not. A decision seemed impossible, given the absence of sufficiently precise probative data. Grant attempted to explain his reasoning at some length. The salesman looked less happy than any clam.

Obviously, the stakes were high – \$30K is a big chunk of change, even on a bloated civil service salary. Grant decided to be really, really careful and think very hard. The salesman interrupted his reverie to ask what he wanted to do. Grant told him he had decided to buy the car, but also not to buy it – to hand over the cheque, but keep it too. The salesman told Grant to leave before he called the police. Grant couldn’t decide whether to stay or go, so did neither... hence, the call for help from his Blackberry. I then had to decide whether to advance him the bail, but all I had to go on were statistics on how many forensic psychologists overall actually showed up for trial having posted bail.

¹ Written with the able assistance of Grant T. Harris who came up with the Subaru example but denies everything asserted about him here, with the exception of his institutional affiliation.

So, how did Grant get himself into all this trouble? Well, he didn't rely on group data, thus depriving himself of the only relevant information available. Surely, the admonition not to make inferences from group data can't mean what it seems to. The difficulty, I have concluded after an epiphanous experience, is that Grant, along with a great many other psychologists, had misinterpreted the phrase "group data."

There is a sense in which one should not draw conclusions about individuals based on "group data" and that's when the data are in fact collected about groups in the first place, not individuals. Strong relationships, for example, between poverty and crime discovered at the neighbourhood level (i.e., all we know is the level of crime and poverty in various particular neighbourhoods) do not allow us to make inferences about the relationship between poverty and criminality at the individual level. To do so is to commit the *ecological fallacy*. It turns out that poverty is a poor predictor of criminality among individuals, despite being a strong correlate at the group or neighbourhood level.

The data on Outbacks were collected on individual cars and were thus not "group data" (although there was a group of cars involved!). It is reasonable therefore to make an inference about one Outback from knowledge of how many similar Outbacks turned out to be lemons. The same thing is true of actuarial instruments designed to predict recidivism where an individual appraisal is based on the known outcome of a group of similar (i.e., obtaining the same actuarial score) offenders. Because the follow-up data upon which actuarial instruments are based were collected on individuals, they can reasonably be applied to predicting the likelihood of recidivism of a new individual offender without fear of committing the ecological fallacy.