Correlation vs. Causation

Correlation is: A mutual relation between two or more things

Causation is: A relationship in which one action or event is the direct consequence of another

Automatically thinking that a certain indicator has an impact on another indicator is a risky assumption. There are many times when variables and indicators have a mutual relationship that it is easily proven mathematically (correlation). Even so, that does not necessarily mean that one thing had an effect on the other.

Many of the indicators studied here are highly correlated, but one cannot automatically assume that one indicator causes another. This concept (causation) is much harder to prove. The data available on www.sumn.org cannot be used to prove causation.

One might argue that heavy drinking of alcohol tends to have a high correlation with more crime. This may or may not be true, and depends on the type of crime. There is a perfect correlation in the case of driving while intoxicated (DWI). Consuming enough alcohol to have a 0.08 blood alcohol concentration and driving a vehicle causes DWI's. A person pulled over for reckless driving who had not consumed any alcohol would not get a DWI—they might however get a ticket for speeding or reckless endangerment. Another form of crime that shows high correlation with heavy alcohol use is violent crime. It has been demonstrated that alcohol use is associated with aggressive behavior. However, aggressive behavior also occurs in the absence of alcohol consumption. Further, not all people who abuse alcohol become violent or aggressive towards others.

You could say: "It is estimated that women whose partners abused alcohol were 3.6 times more likely than other women to be assaulted by their partners."

You can't say: "Alcohol abuse causes domestic violence."

For a second example, we can look at lung cancer. It has been clinically proven that an individual's smoking behavior can cause them to develop lung cancer. It is estimated that 90% of lung cancer deaths among males and 79% of lung cancer deaths among females in the United States are smoking-related. While these two indicators are very highly correlated, people are susceptible to lung cancer other ways (e.g. genetically, 2nd hand smoke exposure, mesothelioma).

You could say: "Cigarette smoking is a risk factor for lung cancer."

You can't say: "If a person develops lung cancer, they must have been a smoker."

Another issue to consider—there might be a third indicator at play (a confounding factor). For example, the number of storks per year nesting in small villages of a given country and the number of newborns in these villages were seen to be associated— the more storks there were, the more newborns per year (this example is attributed to Yule according to Neyman (1952); see also Hofer et al., 2004). Where does the association come from? A closer look reveals that both the number of storks and the number of newborns reflect the size of a village: a larger village has more families producing more newborns and has more roofs allowing more storks to nest.

Another example: the shoe size of grade school students and the student's vocabulary may be highly correlated—the larger the shoe size, the larger the vocabulary the student has. It's plain to see that shoe size and vocabulary have nothing to do with each other, but they are highly correlated. The confounding

factor is age—the older the grade school student the larger the shoe size and the larger the vocabulary.

Demographic and social characteristics—gender, age, race/ethnicity, nationality, county of residence, education level, income level—do not cause substance use or abuse. In 2004/2005 past-month marijuana use was reported by 19% of 18-20 year-old Minnesotans, 14% of 21-24 year-olds, 4% of 25-44 year olds, 2% of 45-64 year-olds, and less than one percent of persons aged 65 and older. While marijuana use is correlated with age, a person's age does not cause their use of marijuana. Seventy-nine percent of 18-20 year-olds did not report use!

Remember: *causation* causes *correlation*. The reverse is not necessarily true (correlation does not prove causation).