

WQD7001 PRINCIPLES OF DATA SCIENCE TUTORIAL 7

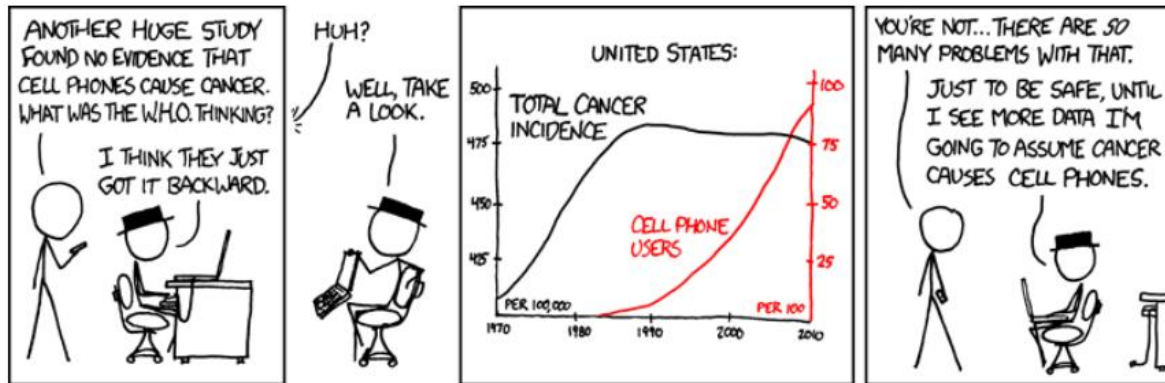


Image Source: <https://xkcd.com/925/>

1. Refer to the cartoon above. Are the assumptions made plausible? Why?
2. Let's say you are screening young patients for an illness that does not manifest until later in life. The illness has no clinical signs or symptoms in the young but can be identified by a laboratory test. The prevalence of the illness in the population is 1/1000. The test has a 100% sensitivity and a 95% specificity. What is the likelihood that a subject with a positive test has the illness?
3. What is population? How does sample differs from population? Give appropriate example for each.
4. What would be possible sources for primary and secondary data?