



Comments on the article :

False-Positive Results in Cancer Epidemiology: A Plea for Epistemological Modesty

Paolo Boffetta , Joseph K . McLaughlin , Carlo La Vecchia , Robert E . Tarone , Loren Lipworth , William J . Blot
J Natl Cancer Inst 2008;100: 988 – 995

by Bernard Beauzamy

April 20, 2009

The authors state quite rightly that : "Chance, which is still an underappreciated problem in our field, seems likely" (to contribute to false-positive results). However, they fail to provide a quantitative estimate about what chance may bring.

Assume that some disease (say : some type of cancer) is just due to chance, and hits everyone with probability p every year (say for instance $p = 1/1000$). Assume that this is just bad luck : people are hit independently (no contagion) and independently one year and the next.

Take many cities of $N = 100,000$ inhabitants. Then, as everyone knows, the average number of diseases per year in each city is $Np = 100$. But the true number is likely to be anywhere in the interval $\left[Np - \sqrt{Np(1-p)} , Np + \sqrt{Np(1-p)} \right]$, that is here the interval 90-110.

If you take a smaller interval, say 95-105, it is normal that a significant proportion of the results will fall outside this smaller interval : In our case, we may expect that 30 % of the results will fall outside this smaller interval.

If you take a very high value, say 300, it is normal that such high values will be recorded, if your observations are numerous enough (many cities and/or many years). The same if you take extremely small values (say 10).

In short, Nature requires heterogeneity, not homogeneity. If a disease is due to pure chance, you cannot expect all the cities with same number of habitants to have the same number of diseases, and you will (sooner or later) find a city with few diseases, and a city with large number of diseases.

A comparison may be made with coin tossing. Assume I play with someone, and I receive 1 Euro if I win and I give 1 Euro when I lose (probability 1/2). Then there will be a time when I have won 1 billion Euros, and there will be a time when I have lost 1 billion Euros. My gain (algebraic : it may be positive or negative) does not tend to 0 (contradicting what most people think) ; it slowly oscillates and will take very high values.