

Ethical Statistics

His name was Merck Smithkline. He was the author on a study of Trioxx, an antiinflammatory drug used to treat arthritis, for which it was known to be effective. The study followed over 500 long-time Trioxx users and an equal number of control subjects who had never used the drug. Dr. Smithkline was looking for correlations between the use of Trioxx and the incidence of any disease other than arthritis, in any demographic group. He noted in the study that African Americans, in the age range from 45-55, male or female, and who had been a vegetarian for more than 6 months at any time in their lives before using Trioxx, had a "strong correlation" between the use of Trioxx and the incidence of restless leg syndrome (RLS). The numbers were as follows:

Total African american former vegetarians age 45-55 : 32
Total African american former vegetarians age 45-55 Trioxx users : 16
Total African american former vegetarians age 45-55 Trioxx non-users : 16
Total African american former vegetarians age 45-55 w/RLS : 4
Total African american former vegetarians age 45-55 w/o RLS : 28
African american former vegetarians age 45-55 Trioxx users w/RLS: 4
African american former vegetarians age 45-55 Trioxx users w/o RLS: 12
African american former vegetarians age 45-55 Trioxx non-users w/RLS: 0
African american former vegetarians age 45-55 Trioxx non-users w/o RLS: 16

Dr. Smithkline correctly calculated the correlation between Trioxx and RLS as follows:

$$\text{Corr} = \frac{\sum (u_i - \langle u \rangle)(r_i - \langle r \rangle)}{\sqrt{\sum (u_i - \langle u \rangle)^2 \sum (r_i - \langle r \rangle)^2}}, \text{ where } u_i = 1 \text{ if subject } i \text{ is a user, and 0 otherwise.}$$

Also, r_i is 1 if the subject has RLS and 0 otherwise. The sums were carried out over all 32 subjects in the subset, and the resulting correlation was +0.378

This confidence level was cited as 99%, since the p-value for this correlation was 0.01, The sample size of 32 and the uneven distribution of subjects with RLS were taken into account.

The data itself was collected correctly and the calculations were correct, both for the correlation and its confidence. Yet Merck Smithkline did something unethical in his report. What was it? What questions would you ask him to reveal his dishonesty?