Re: Is this enough information to make an inference?

Source: http://sci.tech−archive.net/Archive/sci.stat.math/2008−01/msg00025.html

- *From*: "michalchik@xxxxxxx" <michalchik@xxxxxxx>
- *Date*: Wed, 2 Jan 2008 14:31:24 −0800 (PST)

On Jan 1, 8:41 pm, David Winsemius <doe_s...@xxxxxxxxxxx> wrote:

> "michalc...@xxxxxxx" <michalc...@xxxxxxx> wrote
> innews:3b54f97d−f031−420e−a306−6c1fee27a366@xxxxxxxxxxxxxxxxxxxxxxxxxxx:
>
> > On Dec 30 2007, 5:34 pm, David Winsemius <doe_s...@xxxxxxxxxxx>
> > wrote:
> >
> > > "michalc...@xxxxxxx" <michalc...@xxxxxxx> wrote
> > > innews:50af03db−c506−41ef−
> >
> > a96d−cc7a283f2...@xxxxxxxxxxxxxxxxxxxxxxxxxxx:
> >
> > > On Dec 29, 9:04 pm, David Winsemius
> > > <doe_s...@xxxxxxxxxxx> wrote:
> > >
> > > > then you should check the
> > > > spellings underscored (if
> > > > you read USENET with
> > > > monospcaced fonts
> > > > as is the accepted
> > > > convention).
> > >
> > > Sorry about the misspellings. They are
> > > mostly typo's but I should
> > > have used a spell checker. I am not going to
> > > use this as a class
> > > exercise in the foreseeable future. BTW you
> > > misspelled mono−spaced

;−)

First Law of the Interent, eh?

Did you get the statistical perspective to answer your questions. At
one point ISTR that you were asking about attributable risk. I don't
think that the Tomsky invocation of Bayes' Theorem actually gives you
AR. In epidemiology there are several terms that are used to describe
the excess risk for condition X associated with exposure Y. Attributable risk and attributable risk percent are distinct concepts. Are you still looking for something along those lines?

––
David Winsemius

Yep!

I don't mean to be ungrateful but it kind of amazes me that we are on message 27 and my questions really haven't been answered yet ;−)

You didn't really frame your question in a manner that would make an audience take it particularly seriously. You were asking a question in a technical newsgroup. You presented a very sketchy scientific basis for asking the question. You could have done a bit of googling and pointed the audience to this website for instance:

<http://www.pbrc.edu/About_Us/The_Explorers/Faculty_Bio.asp?EmployeeID...>

When I googled a bit more, I found numbers for viral antibody prevalence among obese individuals of 30% and 11% among non−obese persons. <http://www.ncbi.nlm.nih.gov/pubmed/17908526>

The numbers that were presented (rounded to the nearest 10%) appeared as though they were probably pulled out of the air by a teacher (as it turned out you are, but perhaps they weren't arbitrarily chosen. Nor did

you give us your educational background or reason for the question so that we could construct an appropriate answer. You might want to look at:

<http://catb.org/~esr/faqs/smart−questions.html>

The questions again:
Given:
1) The virus is a causal risk factor for obesity and that obesity is not a risk factor for catching the virus.
(That is not actually "given" but rather the subject of research.)
2) 20% of the general population tests positive for the virus.
3) 40% of obese people test positive for the virus.
4) 30% of the United States is obese.

Can we infer?
1) What percentage of obesity in the general population is attributable to the effect of the virus.
2) What is the likelihood that you will become obese if you contract the virus.
−−−−−−−−−−−−−−−end your original questions−−−−−−−−−−−−−−

Epidemiologists usually set up their data as 2 x 2 tables (as always on USENET, these need to be viewed in monospaced font):

```
          Exposed   | Not−exp   row totals
Disease    |   a    |   b     | a+b
−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−
Not_diseased|   c    |   d     | c+d
−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−
 col totals   a+c        b+d    | N (if counts, 1.0 if proportions)
```

The manner in which these groups are collected is crucial to doing a correct analysis and this was not specified in the original question.

```
          Virus+  | No−virus   row totals
Obese     |   a    |   b     | 0.30  <− from 4)
−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−
Not obese |   c    |   d     | 0.70 (by subtraction)
−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−
 col totals  0.20      0.80            |1.0
        ^^from 2)^     \___by subtraction
```

The confusion exhibited by Afonso regarded how to apply 3).

There is a difference between Pr(A|B) and Pr(A&B). The first expression is the probability of A given that B is true, while the latter is the probability of both A and B being true. You could think of Pr(A|B) as narrowing the consideration only to the B is true population. Under assumptions of independence (which are clearly not applicable when A causes B or vice versa) the "a" in the table above would be 0.06 = (0.3 X 0.2). But you told us that among the 30% of the population which

was obese that 40% of them test positive. So a= 0.4 * 0.3 or 0.12, ... twice what would be expected if "obese" and "virus" were independent (and assuming that the numbers were sufficient to give nice, narrow confidence intervals.) So the final table looks like:

	Virus+	No−virus	row−totals
Obese	0.12	0.18	0.30
Not−obese	0.08	0.62	0.70 (by subtraction)
col−totals	0.20	0.80	1.0

Question 1: One way of thinking about "attributable risk" is the excess risk in the exposed groups above what would be expected if there were no virus. The prevalence of obesity in the non−viral exposed group is 0.08/0.7 or 0.114. Prevalence of obesity in the viral groups is ...40% as given, ... so the excess risk could be calculated as 40%−11.4% or 28.6%.

There are other ways for expressing the risk after exposure. This next idea follows what is called the etiologic fraction by some. You could calculate an attributable risk percent (which might be the prevalence of exposure (0.20) times excess risk divided by prevalence of obesity among the viral exposed. 0.2*((28.6%)/40%) , ... or perhaps you want a risk ratio. Since you have not really told us what you are looking for, I am not going to list all the effect measures and their definitions.

Question 2 seems a bit ambiguous to me. You could just want the 40% fraction that you offered. Or you could want the AR or the AR% that was calculated above. Or perhaps you wanted Tomsky's Pr(obese&virus) calculation (which is the same as my "a" above.)

−−
David Winsemius− Hide quoted text −

− Show quoted text −

Thanks for the detailed reply. I will take your criticisms to heart.
Mea Culpa.
.