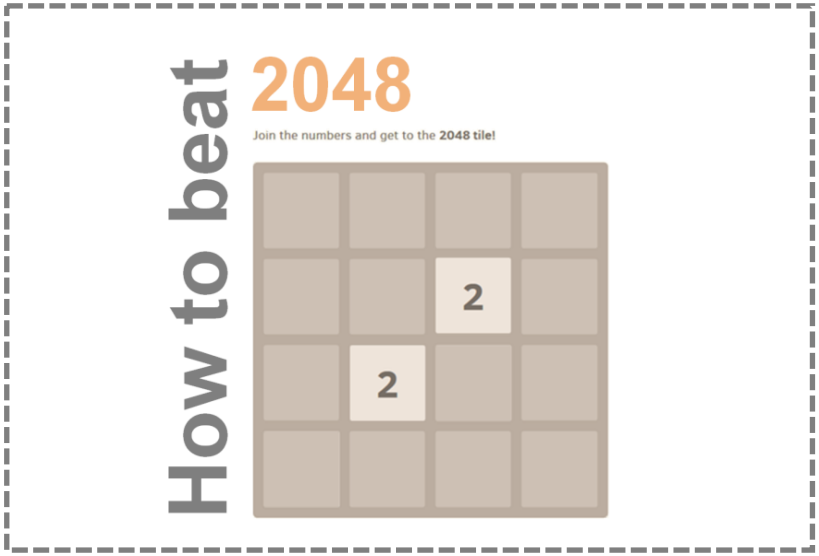
HOW TO USE  DATA TO GET YOUR WAY

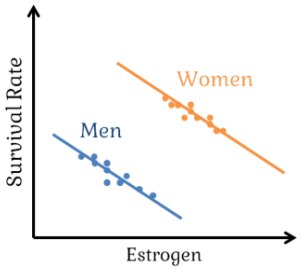
On the last episode…



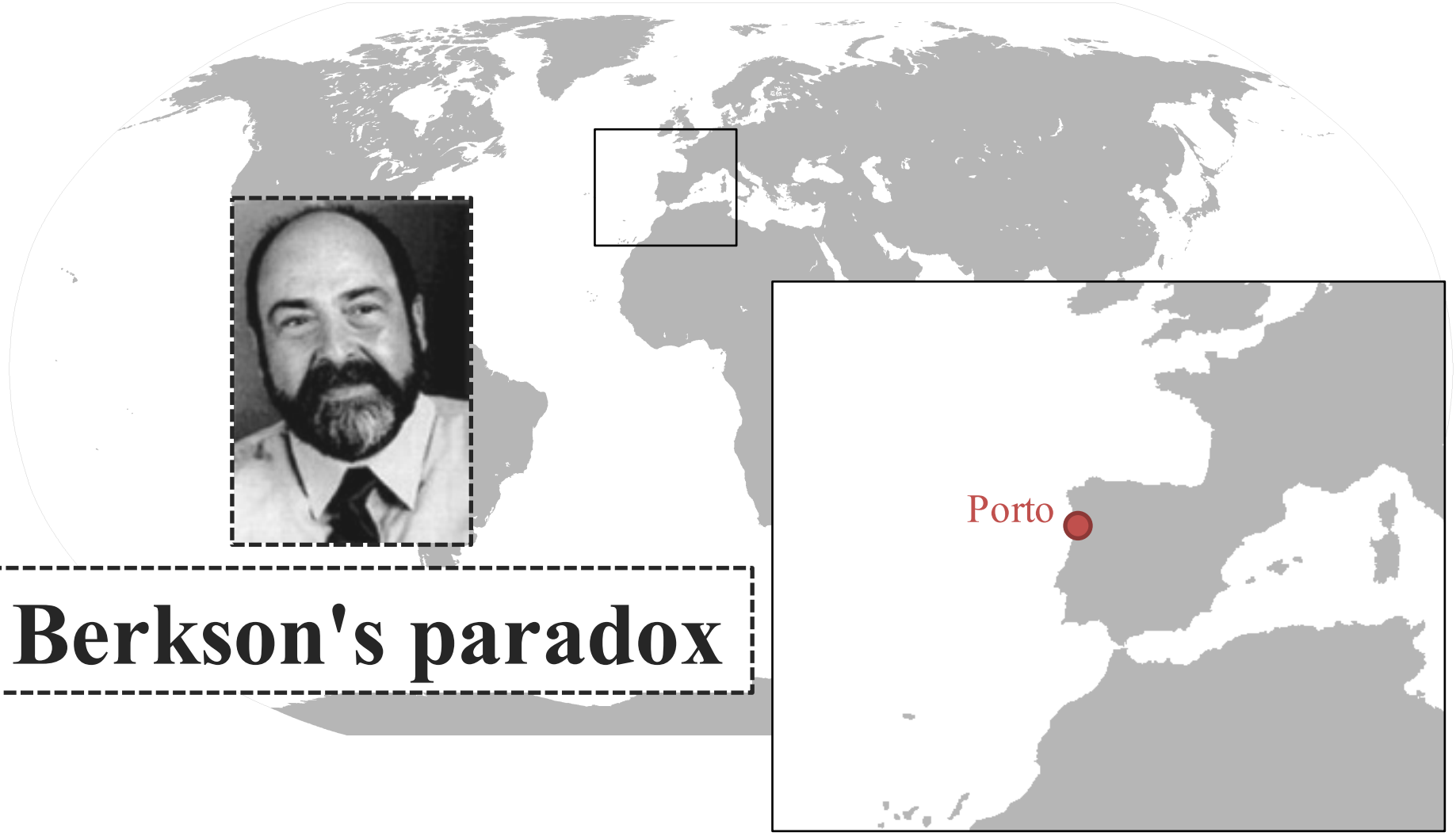
Simpson's paradox 3/3

Say you are running a drug trial and for 2 weeks you go around, you grab people that are sick and you test 2 different vaccines:

	Vaccine A	Vaccine B
Week 1	45/65 69%	25/35 **71%**
Week 2	15/35 43%	30/65 **46%**
Total	60/100 **60%**	55/100 55%

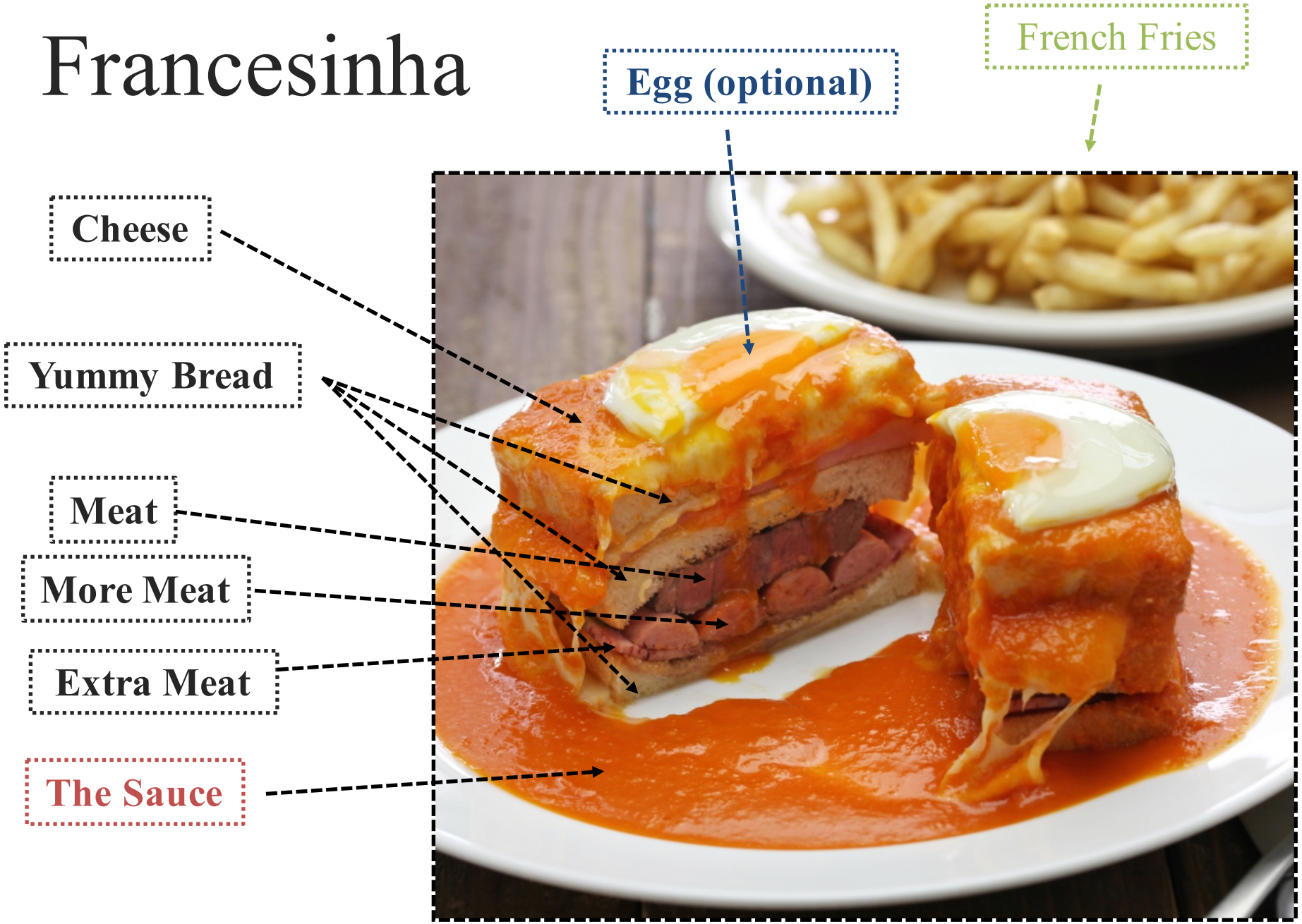


Today…



Berkson's paradox

Francesinha



The Argument

Good Sauce

$\oplus$

Good Sandwich

Good Sandwich

$\updownarrow$

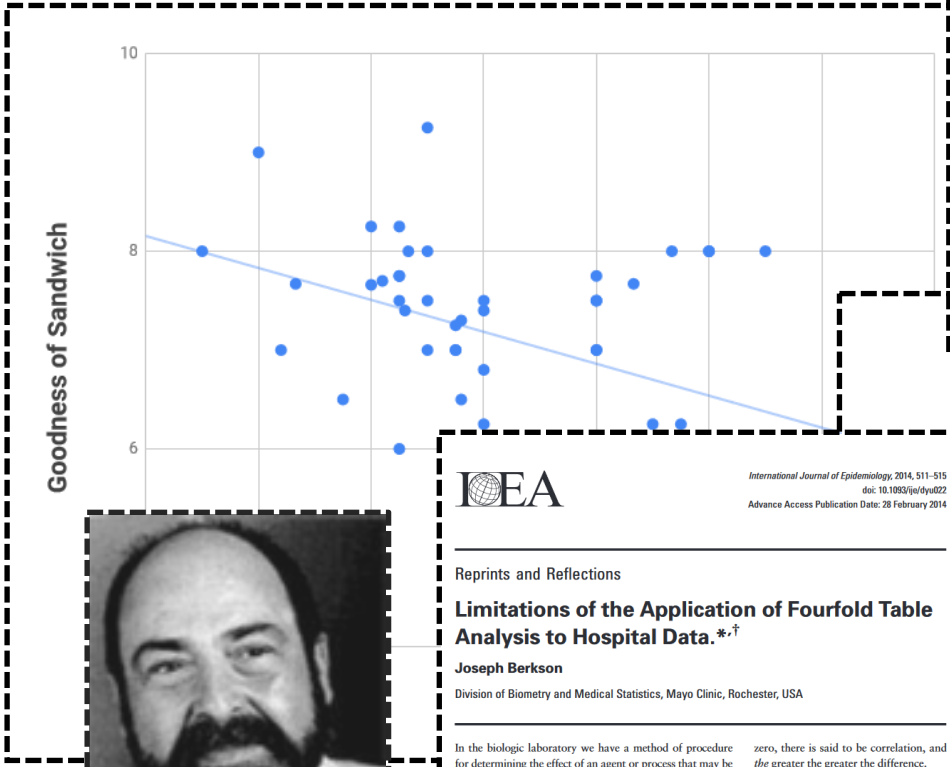
Good Sauce

The Data

University Teacher

Francesinha Reviews

The Problem



IJEA

International Journal of Epidemiology, 2014, 511–515
doi: 10.1093/ije/dyu022
Advance Access Publication Date: 28 February 2014

Reprints and Reflections

Limitations of the Application of Fourfold Table Analysis to Hospital Data.*,†

Joseph Berkson

Division of Biometry and Medical Statistics, Mayo Clinic, Rochester, USA

In the biologic laboratory we have a method of procedure for determining the effect of an agent or process that may be considered typical. It consists in dividing a group of animals into two cohorts, one considered the "experimental group," the other the "control." On the experimental group some variable is brought to play; the control is left alone. The results are set up as in table 1-a. If the results show that the ratio $a:a+b$ is different from the ratio $c:c+d$, it is considered demonstrated that the process brought to bear on the experimental group has had a significant effect.

A similar method is prevalent in statistical practice, which I venture to think has come into authority because of its apparent equivalence to the experimental procedure. In Biometrika it is referred to as the fourfold table and it is used as a paradigm of statistical analysis. The usual arrangement is that given in table 1-b. The entries, a, b, c and d are manipulated arithmetically to determine whether there is any correlation between A and B. A considerable zero, there is said to be correlation, and the correlation is *the* greater the greater the difference.

Now there is a distinction between the method as used in the laboratory and as applied in practical statistics. In the experimental situation, the groups, B and not B, are selected *before* the subgroupings, A and not A, are effected; that is, we start with a total group of unaffected animals. In the statistical application, the groupings, B and not B, are made *after* the subgroupings, A and not A, are already determined; that is, all the effects are already produced *before* the investigation starts. In the end, the tables of the results which are drawn up *look* alike for the two cases, but they have been arrived at differently. Correlative to this difference, a different interpretation may apply to the results, and this paper deals with a specific case of a kind that arises frequently in a medical clinic or a hospital. I take an example.

There was prevalent an impression that cholecystic disease is a provocative agent in the causation or aggravation

The Solution

Get more samples

Thank you

tiagovieira.pt

email@tiagovieira.pt

@TVieira

TMMV

tiagommvieira