Speculative Statistics and Public Policy

Milo Schield
W. M. Keck Statistical Literacy Project
Augsburg College

National Numeracy Network Conference
U. Wash. Boethel, WA 2 May 2009
www.StatLit.org/pdf/2009NNN6up.pdf

Certified Deaths: Observed causes

Death is a fact, but there two kinds of causes:
- Coroner-certified observed causes
- Statistically-attributed inferred causes.

Examples of **coroner-certified causes** and numbers:
- Natural: heart 750K, cancer 556K, alcohol 17K
- Accidental: traffic 45K
- Other: homicide 57K, suicide
- Firearms (30K)

Speculative Statistics

Statistically-inferred (attributed) causes:
- 435K – smoking
- 160K – eating meat
- 75K – gap in quality healthcare
- 70K – pollution-related
- 50K – second-hand smoke
- 22K – radon
- 5K – soot pollution

Inferred Deaths due to Obesity

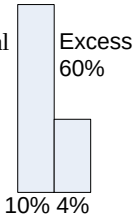
Deaths **attributed** without asserting causation.

CDC: Deaths **attributed to** overweight/obese:
- 2004 estimate: 400,000 deaths/year
- 2007 estimate: 26,000 deaths/year.

To see why, first review two ideas:
- **Percentage attributed**
- **Cases attributed**

1a: Percentage of Deaths Attributed to City Hospital

Suppose these are the death rates for two hospitals:
- 10% for the city research hospital
- 4% for the rural non-research hospital



Based on this alone, one can say:
- "60% of the deaths at city hospital are attributed to that hospital."

The math is simple:
- Excess / Exposed = (10% - 4%)/10% = 60%

1b: Number of Deaths Attributed to City Hospital

Suppose the hospitals have this many patients:
- 100,000 at the city research hospital
- 10,000 at the rural non-research hospital

Number of deaths (given 10% and 4% death rates):
- 10,000 deaths at city research hospital
- 400 deaths at rural non-research hospital.

6,000 (60%) of the 10,000 observed deaths at city hospital are attributable to that hospital.

May 2009 NNN 7

Decisions Based on Statistical Deaths

With this data, what should we do?

- Fire the managers at this "hospital of death"?
- Retrain the staff
- Recommend that patients avoid this hospital?
- Stop funding under-performing hospitals?
- Close under-performing hospitals?

May 2009 NNN 8

Association is not Causation

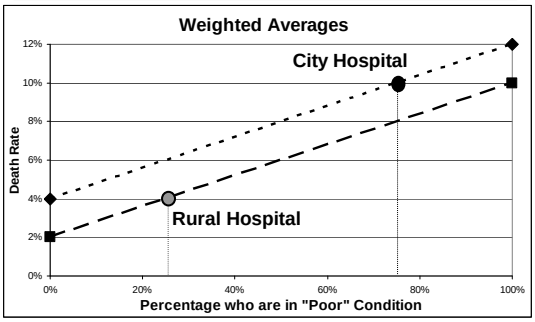
Statistical associations can be due to various influences other than direct internal causation:

- Confounding – influence of associated factors.
- Assembly – the choice of definitions and groups.
- Randomness – most influential in small samples.
- Error/bias – subject, measurement or sampling bias.

Let's examine confounding using a graphical technique.

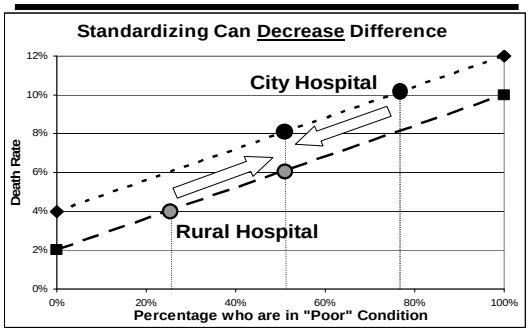
May 2009 NNN 9

Weighted Averages: Graphical View



May 2009 NNN 10

"Take into Account" by Standardizing

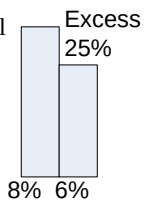


May 2009 NNN 11

2a: Percentage of Deaths Attributed to City Hospital

Consider these confounder-adjusted death rates:

- 8% for the city research hospital
- 6% for the rural non-research hospital



Based on this, one can say

- "25% of these *adjusted deaths* at city hospital are attributed to that hospital."

The math is simple:

- Excess / Exposed = (8% - 6%)/8% = 25%

May 2009 NNN 12

2b: Number of Deaths Attributed to City Hospital

Suppose these are patients at two hospitals:

- 100,000 at the city research hospital
- 10,000 at the rural non-research hospital

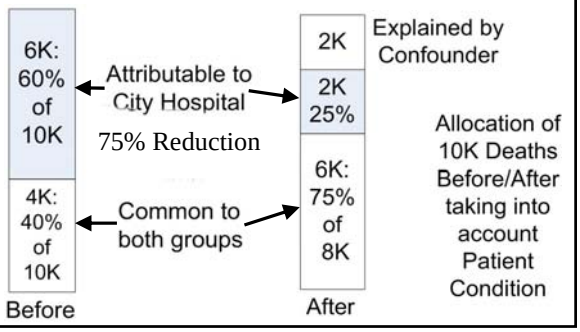
The # of deaths (given 8% and 6% death rates):

- 8,000 deaths at the city research hospital
- 600 deaths at the rural non-research hospital.

2,000 (25%) of the 8,000 deaths at city hospital are attributable to the hospital.

May 2009 NNN 13

Deaths Attributable: Before vs. After Adjust



May 2009 NNN 14

Redo CDC estimate of deaths attributed to overweight

Suppose the annual death rate is:

- 1.6% for overweight; 1.3% for non-overweight.

Excess percentage = 100%*(1.6-1.3)/1.6 = 18%.

18% of overweight deaths attributable to overweight

Suppose: 300 million people, 30% are overweight.
Deaths for overweight: 90M x 1.6% = 1.4 M
1.4M x 18% = 254,000 – close to CDC estimate.

May 2009 NNN 15

Plausible Confounders

What confounders might be positively associated

- with the outcome (death) and
- with the predictor (overweight)?

Taking into account such confounders will decrease the observed association between overweight and death.

Candidates:

- Health
- Heredity
- Occupation (risk taking)

May 2009 NNN 16

Associations with Obesity Obese: BMI > 30

In 2007, US adult obesity (25.6%) was higher among:

- men (**26.4%**) than women (24.8%)
- just high school grads (**29%**) than college (20%)
- blacks (**36%**) than whites (24.5%).
- diabetics (**55%**) [in 2002]

http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5728a1.htm?s_cid=mm5728a1_e#tab

Age	18-29	30-39	40-49	50-59	60-69	70+
Obese	19.1%	26.5%	27.8%	**30.9%**	29.9%	19.4%

Obesity: BMI over 30 (based on self-report)

May 2009 NNN 17

Associations with Death

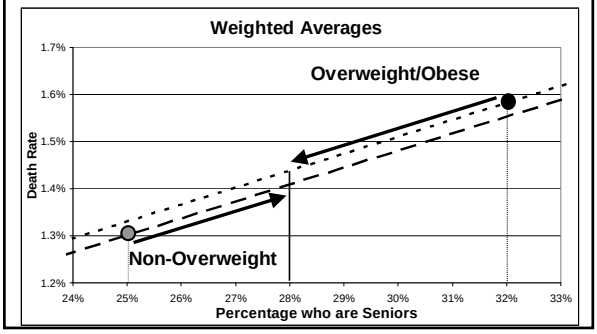
Suppose we had these relative risks for death:

- 1.1: men vs. women
- 1.2: high school grads vs. college grads
- 1.3: blacks vs. whites
- 2.5: diabetics vs. general population
- 10: seniors (over 50) vs. non-seniors.

Age may not be strongly associated with obesity, but it is very strongly associated with death.
Conclusion: Age should be taken into account.

May 2009 NNN 18

Suppose we have these death rates by weight and age



Take into account the influence of old age!!! Duh

Suppose: 40% of population are over 50.

Given this data, the adjusted death rates would be:

- 1.44% for overweight; 1.41% non-overweight.

Excess % = 100%*(1.44-1.41)/1.44 = 2.1%.
2.1% of overweight deaths **attributed to** overweight

Suppose: 300 million people, 25% are obese.
Deaths **due to** overweight: 90M x 1.44% = 1.3 M
1.3M x 2.1% = 27,000 – close to CDC adjusted estimate

Speculative Statistics in the News

Speculative statistics – epidemiologically-based statistics – **are common – but hidden** – in the news.

- No unique grammar or keywords
- Look plausible -- things coroners or hospitals count
- *Journalists don't question them*

*Example: "**13% of the cancers** of the breast, liver, rectum and upper respiratory /gastrointestinal system found among women may **be related to alcohol use**."*
Healthy Years, May 2009. UCLA Dept of Geriatrics.

Speculative Statistics and Public Policy

Policy makers need statistics to make good decision. But to make good decisions, they must be able to distinguish five different kinds of statistics:

- Descriptive statistics (census)
- Descriptive statistics (sample-based)
- Inferential statistics: Margin of Error, Stat. Sig.
- *Predictive (modeled) statistics*
- *Speculative (epidemiological) statistics*

The last two are most difficult to identify.

Conclusion

To help people analyze numbers in the news, Statistical Literacy must focus on:

- the influence of confounding on associations
- Specifically, the influence of confounding on events attributable to an associated factor.

Students are amazed that "attributed to" or "due to" is totally unrelated to any causal claim.

Students are dismayed to learn that these numbers are so soft – so easily influenced by other factors.