Confound those Speculative Statistics!

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

American Statistical Association
Washington DC 5 August 2009
www.StatLit.org/pdf/2009SchieldASA6up.pdf

Let's Talk About A Big Subject

Over Weight

Certified Deaths: Two Kinds of Causes

Death is a fact, but there two kinds of causes:

1. **Coroner-certified (observed) causes**
   - **Natural**: heart (861k), cancer (554k)
   - **Accidental**: traffic (45K), falls (19k)
   - **Other**: suicide (32k), alcohol (17k)

2. **Statistically-linked (attributed) causes.**

Statistically-linked 'Speculative' Statistics

Statistically-linked (attributed) deaths*:
- **435k – smoking (primary smoke)**
- **400k – overweight (CDC, 2003)**
- 160k – eating meat
- 75k – gap in quality healthcare
- 70k – pollution-related
- 50k – second-hand smoke
- 22k – radon
- 5k – soot pollution

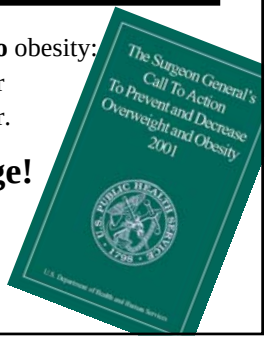
* Web sources: not necessarily reliable.

Speculative Deaths: Deaths Due to Obesity

CDC: Deaths **attributed to** obesity:
- 2003: 400,000 deaths/yr
- 2004: 26,000 deaths/yr.

This is a big change!

To see how, consider:
- **Percentage due to ...**
- **Number due to ...**



Percentage of Deaths Due to Overweight

Suppose these are the death rates:
- 1.6% for those who are overweight
- 1.2% for those who are not overweight

Based on this alone, one can say:
- "25% of the deaths among the overweight are due to [being] overweight."

The math is simple:
- 25% = Excess/Larger = (1.6% - 1.2%)/1.6%

Number of Deaths Due to Overweight

US adult population: 220 million (73% of 300M).
- ~120 M are overweight (60% of adults)

Actual number of deaths among adults
- 1.65 M overweight deaths (1.6% of 120M)
- 1.30 M non-overweight deaths (1.2% of 100M)

~400,000 deaths (25% of the 1.65 million deaths among overweight adults) are "due to" overweight.

Plausible Confounders

What confounders might be strongly associated with the outcome (death) and the predictor (overweight)?

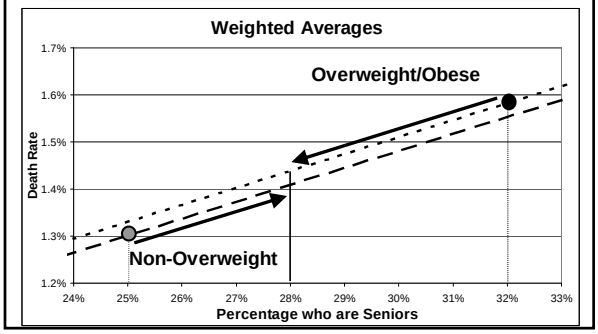
Taking into account confounders may decrease the observed association between overweight and death.

What factors weren't taken into account?
- Diet, exercise and occupation
- Health, heredity, environment

We need a big factor. What could be bigger?

Let's take into account the influence of AGE.

Suppose we have this model by weight and age



Age-Adjusted % of Deaths Due to Overweight

Suppose these are the age-adjusted death rates:
- 1.43% for those who are overweight
- 1.41% for those who are not overweight

Based on this alone, one can say:
- "1.4% of the deaths among the overweight are due to [being] overweight."

The math is simple:
- 1.4% = Excess/Larger = (1.43% - 1.41%)/1.43%

Age-Adjust # of Deaths Due to Overweight

US adult population: 220 million (73% of 300M).
- ~120 M are overweight (60% of adults)

Actual number of deaths among adults:
- 1.72 M overweight deaths (1.43% of 120M)
- 1.41 M non-overweight deaths (1.41% of 100M)

~25,000 deaths (1.4% of the 1.72 M among overweight adults) are "due to" overweight.

Review

Taking into account age changed the number of deaths attributable to (due to) overweight

- from 410,000 (25% of 1.65 M)
- to 25,000 (1.4% of 1.72 M) .

Moral: A small change in assumptions can make a big change in the statistics!

Why?

Why didn't the CDC take into account other factors in their original study?

One plausible explanation is money ($$$)

- **3/9/2004 — The CDC *attributes 400,000 deaths*** to poor diet and physical inactivity. CDC Director Julie Gerberding is a co-author.
- **3/31/ 2004 — CDC director** Julie Gerberding **requests $6.9 billion** from Congress for the agency's 2005 budget...

Speculative Statistics are a Big Problem

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

- No unique grammar or keywords.
- They look plausible -- coroners might count.
- We treat counts as facts.
- Journalists and politicians don't question them.
- We don't question them.

Conclusion?

Public policy (passing laws, taking drugs off the market, demonizing overweight or second hand smoke) is based on these "speculative statistics."

But are these numbers real – or spurious?

Educated adults don't know! They can't tell!

Can traditional statistics address these questions?

If not traditional statistics, then what?

If not statistical educators, then who?

Bibliography

Actual Causes of Death in the United States, 2000. Mokdad, Marks, Stroup and Gerberding. JAMA, March 10, 2004; 291: 1238 - 1245.

Correction: Actual Causes of Death in the United States, 2000. Mokdad, Marks, Stroup, and Gerberding. JAMA. 2005; 293(3):293-294.

Excess Deaths Associated With Underweight, Overweight, and Obesity. Flegal, Graubard, Williamson and Gail. JAMA, April 20, 2005; 293: 1861 - 1867.

An Epidemic of Obesity Myths. The Center for Consumer Freedom. Washington DC.