Introduction: Not everything you read is true

David Schlesinger, Ph.D.

Lars Leksell Gamma Knife Center
University of Virginia

Research support: Elekta Instrument, AB

One year ago...

Statistical Failings That Keep Us All in the Dark

David Schlesinger, Ph.D.
University of Virginia Health Systems, Charlottesville, VA

In the meantime...

The truth is hard to come by

Session Educational objectives

1. Learn about the presence of statistical problems in published studies
2. Identify common signs and symptoms of potential problems in various types of statistical tests
3. Learn methods for correctly implementing statistical analyses of the type commonly found in clinical publications
Cellphone use causes cancer

Major Cell Phone Radiation Study Reignites Cancer Questions

Exposure to radio-frequency radiation linked to tumor formation in rats.

Led by the National Toxicology Program (NTP) under the NIH
Rodents exposed to calibrated RF (GSM and CDMA) radiation for 9 hours/day over 2 years
Division into groups by SAR exposure

Association between exposure and cardiac schwannoma in male rodents (no association in female rodents)


SRS for lung cancer: Does morning or afternoon make a difference?

Human studies are mostly one-sided

<table>
<thead>
<tr>
<th>Publication Year</th>
<th>Study Type</th>
<th># Participants</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>Interphone Study Group</td>
<td>5000 cases; 5000 matched controls; 13 countries</td>
<td>No overall risk*</td>
</tr>
<tr>
<td>2007 (updated 2011)</td>
<td>Danish cohort study</td>
<td>508,000</td>
<td>No association</td>
</tr>
<tr>
<td>2010 (updated 2011)</td>
<td>Million Women Study</td>
<td>417,710</td>
<td>Yes (accidental, misleading); no association</td>
</tr>
<tr>
<td>2013 (updated 2014)</td>
<td>CERENAT</td>
<td>424 cases; 424 matched controls</td>
<td>No association with regular use; yes association with heaviest use</td>
</tr>
<tr>
<td>2014</td>
<td>Prospective cohort study</td>
<td>151 cases; 2438 controls</td>
<td>Increased risk of glioma</td>
</tr>
</tbody>
</table>

Study was released before complete peer review on a pre-publication website
Control rats showed less than expected natural rate of tumor incidence and died early
Incidence of tumor development correlates with age, so the early control death may have magnified the statistical findings

http://arstechnica.com/science/2016/05/study-that-found-cell-phones-cause-cancer-in-rats-is-riddled-with-red-flags/

The result of studies of thousands of animals and hundreds of thousands of people, supported by millions of dollars in funding, is that we have no definitive answer to the question of cellphone use and cancer.

So….

How confident can we be about studies like this:

SRS for lung cancer: Does morning or afternoon make a difference?

Morning vaccination enhances antibody response over afternoon vaccination: A cluster-randomised trial

D. Rahn, et al., Cancer 177(2), 2011.
Radiation Oncology is full of similar studies

Reporting statistical tests has become a requirement

Medicine increasingly relies on p-values

p-values just below p=0.05 are over-represented

Conclusions: p-values immediately below 0.05 appear to be over-represented in the literature relative to their expected frequency.

shows evidence of systematic error including publication bias, selective reporting, methodological errors, or fraud.

Try it yourself: there are many ways to achieve a desired story

The ASA’s statement on p-values: context, process, and purpose

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.
What we think is the truth often can't be replicated

One study's result is not necessarily the truth

http://news.harvard.edu/gazette/story/2016/03/study-that-undercut-psych-research-got-it-wrong/
http://projects.iq.harvard.edu/psychology-replications/

The number of retractions is sharply rising

http://fivethirtyeight.com/features/science-isnt-broken/#part2

A lack of statistical fluency may be part of the problem

Many medical physicists receive little training in practical statistics as applied to clinical outcomes studies. However...these studies are at the heart of our profession. How to recognize when the statistics don't quite add up?
But...we can learn to be better