Toys, Tales, and a Journalist's View of Statistics

Lewis Cope
Minneapolis Star and Tribune (Retired), Minneapolis, MN

Abstract

I tell some tall-tales in talking about statistics. I minimize the math. I oversimplify. And I use my grandson's toys to make my points. OK, I'm just a journalist. But I'm one who gives lectures to fellow journalists about using numbers, because statistics are extremely important in telling our readers and viewers about science, politics, and other items in the news. So this old science writer has a few ideas that just might help promote statistical literacy for the lay public. I tell journalists to use four tests in assessing all scientific studies and opinion polls: check the numbers, while knowing what really counts; look at study design, while remembering the "Rear View Mirror Rule;" look for "other explanations," where one question fits all; and recognize the value of journalistic and peer review. I also talk about missing numbers---a problem in many areas of the news. I use newsy examples, talk about rating risks, and warn about the "Lake Wobegon Effect," based on statistics from that fabled Minnesota community. And my bottom line: Your noggin is even more important than the numbers.

Background: Based on concepts in the book ***News & Numbers*** [2nd edition, by Victor Cohn and Lewis Cope, Blackwell Publishing, ©2001 Cohn and Cope], and from other sources.

1. Introduction

I confess:

I often oversimplify.

I shun most math. I even use some of my grandson's toys to make my points. That's what you get for inviting a journalist to speak to professionals about statistics.

But as a science writer, statistics are a vital part of my job. And I've coauthored a journalism textbook, ***News & Numbers*,**[i] that shows how to use numbers in all areas of news reporting. The book puts an emphasis on conveying statistical concepts to everyday people—the journalists' readers and viewers.

So just maybe, I have a few ideas that may help you promote statistical literacy for the lay public.

As I do when I give lectures to fellow science writers, I'll use a variety of examples and a few obvious tall-tales. I'll alert you to "missing numbers." I'll warn you about the Lake Wobegon Effect. And I'll use the toys (corny as they may be) to leave you with visual reminders of important principles.

So how do we know which scientific studies and other statistical claims to believe? How do we separate the probable truth from the misleading trash? I could suggest dozens of things to consider. You could tell me hundreds. But it's easier to remember a few than many. So I focus on just—

2. Four Key Tests

1. Bigger numbers are better.

> Did you hear about the drug researcher who reported "33 percent were cured, 33 percent died—and the third mouse ran away."
>
> *I unleash my first toy, a mechanical mouse that scoots across the table.*[ii]

It's important to remind our readers and viewers that "statistical significance" isn't some nebulous concept. It simply means that the findings are unlikely to be due to chance alone.

But statistical significance and P values are only the start. Bigger numbers have the power to find an effect when it's there, and the ability to profit from statistical breakdowns. You know all this much better than I. So let's move to—

2. Study design, and the hierarchy of studies.

All studies aren't equal. Some are better than others simply because of how they're designed.

The hierarchy example that everyone gets quickly: A medical study may start in a lab dish, then be done in animals. But you must put it to the test in people before you can claim that it will save lives.

I cite two other study-design features as most important.

> *I look backward with a toy mirror to illustrate why prospective studies are better than retrospective studies*:

Look-back studies are like the limited vision that you get looking through your car's rear-view mirror. Memories fade; records are often incomplete. Forward-looking studies are better.

And in speaking of the Gold Standard of a medical-treatment study, I focus on four phrases:

- The patients are randomly assigned
- to either a treatment or comparison group.
- It's double-blinded and
- (usually) placebo-controlled.

A key aim, of course, is to protect against the "powers of suggestion and expectations."

Now let's get to the big one—

3. Is there another explanation?

I flip a piece of plastic bread and tell this tale:

A father was puzzled and perplexed. Every time that any of his 11 kids dropped a piece of toast on the floor, it landed buttered-side up. "This defies the laws of chance," he said. The "other explanation":

His kids always butter their bread on both sides.[iii]

When we look at anything statistical we need to consider: **Is there another possible explanation?**

With toy chicken in hand, I ask:

Remember the rooster who thought his crowing caused the sun to rise each morning? [iv]

And as I dribble a miniature basketball:

Is it playing a lot of basketball that makes professional basketball players grow so tall? [v]

The rooster and the basketball players remind us:

Association alone doesn't prove causation.

So a virus discovered in a patient's body may not be the cause of an illness. And the chemical found in a workplace may not be the culprit in an outbreak there. More is needed to confirm such links.

A timing example: Autism starts at the age when toddlers get their vaccine-shots. But extensive studies have found no good evidence of any cause-and-effect link.

There are other types of "other explanations." I talk a three-step walk*, as an action-figure toy takes the steps:*

(1) Did the study last **long enough?** Did a climate study continue long enough to pick up real trends, not just the normal ups-and-downs of weather? *In medicine:* Are cures claimed too early? If the study passes this test—

(2) Does the **difference make a difference?** Is a detected bodily change just normal variability? Does an early-detection test find a problem early enough for treatment to actually help? And if the study passes this test—

(3) Should the findings be applied **broadly or narrowly**? *One recent case in the news*: Do the findings from a heart-attack-prevention study really apply to all patients? Or, is the real explanation that the findings apply only to high-risk patients like those in this particular study?

And then I discuss how:

Confounders can certainly confuse.

Dropouts can damage, even destroy an otherwise good study.

And the **Healthy Worker Effect** just happens to be a handy way to explain another important point: Care must be used to make sure any **comparison group is really comparable.**

To my mind, all these things—and the list could go on—are the same in two key ways:

First, all show how easy it is to be misled unless great care is taken. Of course, more figures may help us to figure out what's-what.

Second, a single question can go a long way toward identifying such problems when they exist. When you hear any scientific or other statistical claim, simply think and ask:

Is there another possible explanation?

4. Peer review and journalistic review.

I toss a small red flag:

If you watch NFL football on TV, you know that this is a coach's signal to have the referee review the last play.

We all know about the value of peer-review for scientific studies. But I point journalists to four areas that scientists often consider, yet often need further public airing:

Is there possible bias? *I flash a wad of play-money and ask:* Who paid for the study? This isn't an indictment, but may need exploring. Are there any other indications of possible bias, unintentional or otherwise?

Is there a scientific fit? If findings clash with other evidence in the field, you have more questions to ask.

Is it useful and practical? *Waving play-money again:* How much will a new drug (or whatever) cost? Will it be **so expensive** that it will be impractical to use?

What now? What new studies and other steps lay ahead?

Science is complex, and based on probability without waiting for the ever-elusive proof-certain. So some studies will reach less-than-perfect conclusions, some even downright-wrong conclusions.

But science works wonderfully well because one study leads to another, and mid-course corrections are made as necessary. So at any point along the way there is the

certainty of some uncertainty.

My bottom-line message to fellow journalists in ***News & Numbers:***

We need to say "may" and "evidence indicates'' more often, and use the word "proof" less frequently. And as best we can, we need to tell our readers and viewers the degree of uncertainty about a study's conclusions.

Now let's look at the same things, in another way

3. Probing the Polls

Opinion polls are a major way that people come into contact with conclusions that need to pass statistical muster. And the same four basic tests apply:

1. The people interviewed must be a **random sample** of the population whose views we want to learn about. This makes possible: 1 = 200,000. That is, a single person interviewed can represent up to 200,000 potential voters across the nation. Without random sampling, this doesn't work. Without random sampling, it's not a scientific poll. *This is **study design**.*
2. The more people that are polled, the smaller the "**margin of sampling error**." I say "sampling error," because other errors can creep into polls. "***Bigger numbers are better.***"
3. The poll questions asked must be **clear**, not confusing. They must **not tilt** toward a particular answer. Ask: Does a poll's questions leave open the possibility of ***"other explanations"?***

Here's a recent mushy poll question:

After President Bush's reelection, a poll found that many voters cited "moral" issues as the key reason that they had voted as they did. Folks on the Christian Right then said "I told you so." But folks on the Liberal Left said they voted for <u>their</u> moral issues – such as ending the death penalty, and providing more aid to the poor.

4. Two "red flag" ***journalistic-review*** questions: Who paid for the poll? And what have other recent polls found about the candidates or the subject at issue?

Even when done correctly, a poll is still (*as I click a toy camera*) only a **snapshot,** capturing **one point in time**. Things can change quickly. ***The certainty of some uncertainty***.

Now let's switch to —

4. Missing Numbers

Flashing a toy police badge, I ask:

Do you remember TV's Sgt. Friday? The Dragnet cop's signature line was: "Just give me the facts, ma'am."

To paraphrase Sgt. Friday:

> "Just give me the numbers, ma'am."

Take risk-appraisal as an example. Here, from ***News & Numbers***, are some of the needed numbers that too often are missing in news reports[vi]:

1. Put a **number** on a risk rather than just saying it's large or small. Is it 1 in 1,000, or 1 in a million?
2. Put a **time** on a risk. It's a 1 in 1,000 risk of getting an illness: But is that for one year? Or over the four years of the study? Or over a lifetime?
3. Put a **denominator** on a risk. Thousand of patients had a specific side effect from a new drug; how many people took the drug?
4. **Use ranges,** not just worst-case figures.

Missing numbers come up in other ways:

I've mentioned that extensive studies have found no cause-and-effect link between toddler's immunizations and autism. Still, some worried parents won't allow their children to be vaccinated against measles and other dangerous diseases. In many news reports, the missing numbers are the tolls these childhood diseases took before the vaccines were available.

Another example:

The Bush administration says that it would be too risky for Americans to import prescription drugs from Canada, even though the drugs are available there at much lower costs. The missing numbers here: Vital statistics that show Canadians have a longer life expectancy than we do in the United States.

5. Confuse and Abuse

I also discuss various ways that everyday numbers can confuse and be abused. For example:

You can drown in the middle of a lake that has an average depth of 3 feet. From ***News and Numbers:*** When an industrial plant releases a toxic chemical, the key concern isn't the average amount in the air, but the amount downwind where people are breathing it.

A measurement goes from 2 to 6. To make the increase sound even bigger, someone calls it a 200% increase. But you don't have to, I tell journalists. And I particularly warn them about—

6. The Lake Wobegon Effect

Remember Garrison Keillor's fabled Minnesota town where the women are strong, the men good looking—and *all* the children are above average?[vii]

There are so many ways to measure a good education. No wonder that, across our nation, most schools have found one way or another to prove what a great job they are doing**.**

There are so many health and other statistics that a city or state can find some way to show how safe, healthy or just-plain-great it is. And hospitals, managed-care health plans, corporations, advocacy-groups all do the same.

To journalists: Ask for the numbers that you think you need.

In my last minute, let me mention not what I talk about, but–

7. How I Present It

In addition to **over-simplifying**—

I **limit the lingo**. For example: I mention confounders, but under the umbrella of "other explanations."

I strive for **human terms**. I call the "subjects" in a clinical trial "patients"— or just "people."

I'm kind to people who are **math-challenged**. The only equation I've used is 1 = 200,000, in talking about polling. ***Thinking can be more important than figuring. Your noggin is even more important than the numbers.***

I use **examples from the news** as a reminder of how relevant all this is to all our lives.

Finally, I use my grandson Ethan's toys as **visible reminders.**

And I use the toys to **add a bit of fun.** For all good things should be fun.

Biography

Lewis Cope was a science reporter for the Minneapolis Star Tribune for 29 years, and is a former president of the National Association of Science Writers. He is coauthor of the 2nd edition of ***News & Numbers***, and is a board member of the Council for the Advancement of Science Writing.

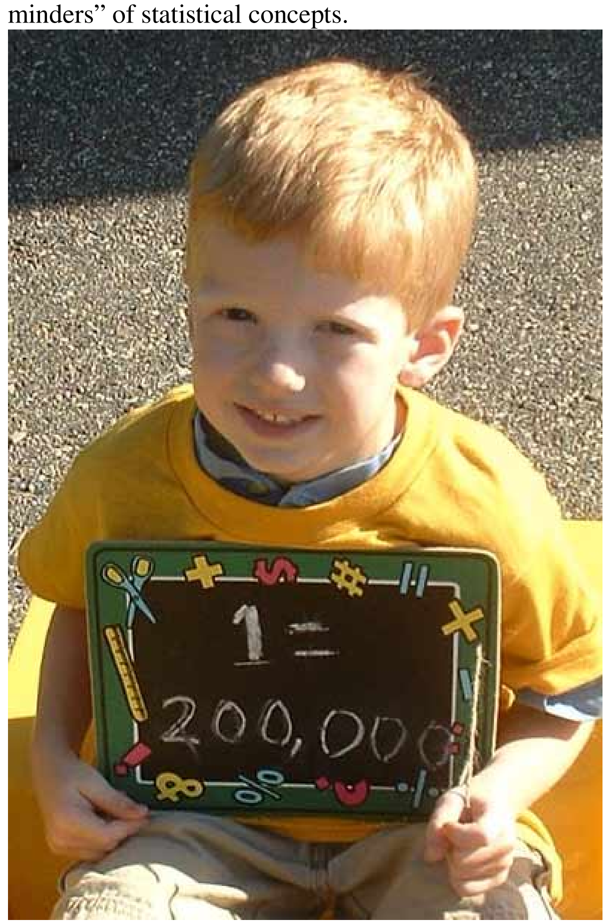
He holds a journalism degree from Washington & Lee University, Lexington Va., and spent a year as a fellow at the Columbia University Graduate School of Journalism, New York City.

Thanks & Endnotes

---

[i] Concepts for this paper were drawn from ***"News & Numbers"*** [2nd edition, by Victor Cohn and Lewis Cope, Blackwell Publishing, ©2001 Cohn and Cope], and from other sources. We didn't use the term "statistical literacy" in the book, but that's the idea behind it: Helping journalists better inform the lay public about all things numbers.

[ii] Thanks go to my 4-year-old grandson, Ethan Bentley. His toy-chest was my inspiration for my "visual reminders" of statistical concepts.



[iii] ***"News & Numbers"*** is rich in various types of examples. The buttered-bread tale is on Page 15.

[iv] The crowing rooster tale, ***"News & Numbers,"*** Page 22.

[v] The tall tale of all tall tales (basketball players), ***"News & Numbers,"*** Page 16. And thanks go to Robert Young, Fox Chase Cancer Center, Philadelphia, for passing it along for the book.

[vi] Eighteen tips on analyzing and writing about risks are on Pages 140-144 in "***News & Numbers."*** Risks are discussed in various ways throughout the book.

[vii] A tip of my cap to Anemona Hartocollis, a *New York Times* writer who drew national attention to the Lake Wobegon Effect: "New Math: No One Is Below Average," *New York Times*, 20 June 1999. And great thanks to Garrison Keillor of public radio, whose magical mind gave us Lake Wobegon.