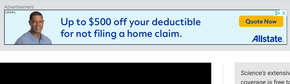
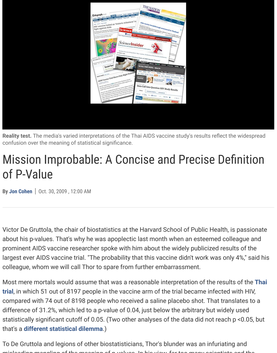
SHARE
13

Read our COVID-19 research and news.

Advertisement




Reality test. The media's varied interpretations of the Thai AIDS vaccine study's results reflect the widespread confusion over the meaning of statistical significance.

Mission Improbable: A Concise and Precise Definition of P-Value

By Jon Cohen | Oct. 30, 2009, 12:00 AM

Victor De Gruttola, the chair of biostatistics at the Harvard School of Public Health, is passionate about his *p*-values. That's why he was apoplectic last month when an esteemed colleague and prominent AIDS vaccine researcher spoke with him about the widely publicized results of the largest ever AIDS vaccine trial. "The probability that this vaccine didn't work was only 4%," said his colleague, whom we will call Thor to spare from further embarrassment.

Most mere mortals would assume that was a reasonable interpretation of the results of the **Thai trial**, in which 51 out of 8197 people in the vaccine arm of the trial became infected with HIV, compared with 74 out of 8198 people who received a saline placebo shot. That translates to a difference of 31.2%, which led to a *p*-value of 0.04, just below the arbitrary but widely used statistically *significant* cutoff of 0.05. (Two other analyses of the data did not reach *p* <0.05, but that's a **different statistical dilemma**.)

To De Gruttola and legions of other biostatisticians, Thor's blunder was an infuriating and misleading mangling of the meaning of p-values. In his view, far too many scientists and the journalists who cover them (**mea culpa**) garble the meaning of the term.

SIGN UP FOR OUR DAILY NEWSLETTER
Get more great content like this delivered right to you!
Email Address *

The debate might seem like a semantic fine point, further evidence that statisticians are from Mars and the rest of us are from Earth. But to De Gruttola and his fellow martians, the widespread confusion about such a fundamental concept has huge stakes: p-values help biomedical researchers determine which products to move forward–or at least, in the case of the Thai study, to build on–and which ones to toss in the trash.

To heal wounds and improve communications between biostatisticians and the confused masses that rely on them, De Gruttola agreed to discuss the details of what p-value means and does not mean with *Science*NOW. But, as you'll see, the probability that this will solve the problem is low.

SN: What does a p-value of 0.05 mean?

V.D.G.: Everything starts with a scientific context in which you're developing an intervention to do something that's going to make people healthier or happier. So there's a goal for this new intervention to have a measurable impact on somebody's life. That goal must be turned into a hypothesis. Often in clinical research, the hypothesis you're testing is that the treatment does not work. One hopes to reject the null hypothesis–namely, that the intervention does not make people happier or healthier–and prove that the treatment in fact does work.

SN: A lot of smart people read *Science*. The results say the p-value is 0.04. How would you say that in a sentence? The journalistic way to handle it is to say there's less than a 5% chance that the results are misleading. But that makes you guys bristle.

V.D.G.: I don't really like that too much. If there's a p-value of 0.04, it says the probability is only 0.04, or one in 25, that you would have seen results in the magnitude that you saw or even larger if there were not real vaccine effects. *Science* readers should know, I believe, what a p-value is.

SN: I can put a finer point on it. I think it means there's a less than 5% chance that I'm going to say this worked when it really didn't.

V.D.G.: It's not that. It's really not like that. It's the difference between I own the house or the house owns me. It's two different concepts. If you're trying to get me to recant like Thomas Moore in the age of Henry the VIII, you're going to have to have a great big ax. I'm not going to recant.

SN: But the average person thinks you did a study, and you want to make sure you're not making a mistake. Something could happen by chance and you could assume something that is real when it isn't. And by 'real' I mean something that would not be reproducible if you were to do it again.

V.D.G.: Reproducibility is a great argument, and I think that's an excellent point. The other way you could say this is assume the vaccine does nothing, it's water. If we did the experiment 25 times, we would expect one in 25 times to get results that are as large or larger than the ones we saw.

SN: What's the phrase you read in the newspaper or hear on the morning news that makes steam come out of your ears?

V.D.G.: Probably the most frustrating one is it's the probability that the vaccine works, which is a different concept. It would be better to say you did the experiment 25 times, and the vaccine really doesn't work, but one in 25 you get results as impressive as the ones you saw. Isn't that going to communicate to your audience?

SN: It is. But it's a lot of words.

V.D.G.: If you have a little insert, you should be able to say that much.

SN: You know the reality is that an editor will look at this and go, wait a second, the plain English here is the *p*-value cut off of 5% means we have 95% confidence that this is real. Does that sit okay with you?

V.D.G.: No! No! That it's real? It's not the probability that the vaccine works. It depends on what "this is real" means. You're saying, "Why can't you say it simply?" You see that language all the time, and clinicians interpret it wrongly. That's like Thor.

SN: I understand that it gets under your skin, but it doesn't seem wrong to me. It seems like poetic license.

V.D.G.: You're saying it's a poetic difference? I say no, it's a huge qualitative difference between what this study shows and the statement that we think the probability that the vaccine works is 96%.

SN: We're at the nub of it now. I know it's pissing you off.

V.D.G.: Right.

SN: To the average person, the probability the vaccine worked one-third of the time is what the study says. In all probability, this vaccine works one out of three times.

V.D.G.: I think that's a great point. So we could look at the confidence interval on triple therapy with antiretrovirals compared to dual therapy. [This was the breakthrough in 1996 that led to the first powerful anti-HIV treatments **ACTG 320** was one of the first studies that showed this.]

SN: I remember when I first saw the data. I said, "Oh, that's what it looks like when antiretrovirals work." I looked at one slide and said, "I've never seen that before." What was the confidence interval?

V.D.G.: I have the data here. The endpoint was AIDS or death, and you had half as many people who had AIDS or death in the good arm versus the bad. The confidence interval was.33 to 0.76. The p-value was 0.001.

SN: So even in something that really works, that's still saying it reduced AIDS or deaths somewhere between 33% to 76%.

V.D.G.: There's still a certain amount of variability, you're right. The main thing that this reflects is sampling variability. That 0.33 and 0.76 means if you replicate this 20 times, 19 times out of 20 on average that interval would contain the truth.

SN: People say to me all the time: 'If you need a statistician to know whether it matters, then it doesn't matter.'

V.D.G.: That's where they're wrong. Take the example of antiretroviral drugs, where you got small effects with studies of individual drugs compared to two drug therapies that you really could not have detected without the right kind of study and right kind of statistics. And then we built up to a more potent effect when we combined three drugs. Sometimes you do make a leap that does not require this type of careful study because it's not incremental. But incremental improvements are often most efficiently investigated by using statistical methods.

SN: Switching gears, you have called the 0.05 cutoff a "fetish."

V.D.G.: It's a lottery idea. One number off and you go from multimillionaire to pauper. It's a perfectly arbitrary thing. You're separating the saved from the damned based on a number. What makes the 95% confidence interval a useful idea is that it's made fairly consistent across studies so you can see a confidence interval and it has a similar meaning.

SN: But in the real world, if you reach statistical significance of 0.05 in your study, your product moves forward.

V.D.G.: Well, the U.S. Food and Drug Administration requires two studies that reach 0.05. And two studies at 0.05 is a much higher level of protection if those are two independent studies. It's 0.05 times 0.05, which is 0.0025. That's why they require two. One study at .05 doesn't mean much. If you're doing a lot of studies, the probability that you'll have one that reaches 0.05 can be quite high.

This is the other point that needs to be taken into consideration–the multiple comparisons. If this were the only vaccine study that had ever been done, you would interpret it differently than if there'd been a number of studies, including studies with the same two components, that were negative.

SN: You're suggesting combining the two earlier AIDS vaccine efficacy studies that have been done?

V.D.G.: You look at the totality of the evidence. For vaccine efficacy, it really doesn't look so good. If this were a p-value of 0.001, like ACTG 320, that's impressive even in the context of a bunch of studies going on, some positive, some negative. The 0.04 in the context of other studies is not really impressive. The probability that one of these studies would give a 0.04 result is 11.3%. It's roughly comparable to the chance of getting snake eyes if you threw the dice three times.

If you do a lot of studies and a bunch of them are positive, then you get very, very excited something is going on. If you do a bunch of studies and only one of them is positive, then you don't get so excited.

SN: That's precisely the opposite of what happens. In the real world, a field like the AIDS vaccine one has so much failure that when researchers see positive data they say, Eureka!

V.D.G.: That's a good point. But you have to take into account the context. Eventually if you shoot craps, you're going to shoot a seven or 11. You could say, "Gee, I threw so many times and didn't get a winner, but then I prayed to my local deity and got a seven. First I prayed to this saint and then another one and then to a third one and finally I prayed to the local one and got a seven." That's the saint that worked for you. That's how humans have reasoned for time immemorial, until people started to get statistical–and started fighting with people like you.

SN: If the AIDS vaccine field has a positive finding, what does it have to be before people can say, "We can take this to the bank, it's real, it's good."

V.D.G.: You want to see a nice, healthy, strong p-value, at least less than 0.01. Then you really start talking about something you want to put your money on.

If you are a glutton for punishment and want to further explore the definition of p-value, peruse the **Wikipedia** *page on the subject. It cannot be proved with any degree of certainty that statisticians actually wrote the page, but in all probability, they did. The definition illustrates why the concept makes even very smart people turn their liquor cabinets: "In statistical hypothesis testing, the p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the hypothesis is true." And the Wiki page tellingly spends more space on "frequent misunderstandings" than it does on the definition itself.*

Posted in: **Math**

Jon Cohen
Jon is a staff writer for *Science*.
Email Jon | Twitter

More from News

Founder of geometric analysis honored with Abel Prize

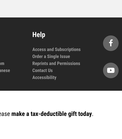
This dizzying labyrinth will host next year's party for maths 'Nobel' prize

Skepticism surrounds renowned mathematician's attempted proof of 160-year-old hypothesis

Science's extensive COVID-19 coverage is free for all readers. To support our nonprofit science journalism, please **make a tax-deductible gift today.**

Got a tip? How to contact the news team

Advertisement


Advertisement


Related Jobs

Investigator Scientist
MRC Laboratory of Molecular Biology
Cambridgeshire

Research Microbiologist (Postdoctoral Research Associate)
USDA ARS-ERRC
Wyndmoor, Pennsylvania

Principal Investigator / Faculty
San Diego Biomedical Research Institute
San Diego, California

MORE JOBS ▶

*Science*Insider

'We've got to be able to move more quickly.' The pandemic reality of COVID-19 clinical trials
BY JENNIFER COUZIN-FRANKEL | JUN. 16, 2020

Huge open-access journal deal inked by University of California and Springer Nature
BY JEFFREY BRAINARD | JUN. 16, 2020

HIV and TB increase death risk from COVID-19, study finds–but not by much
BY LINDA NORDLING | JUN. 15, 2020

FDA just gave a thumbs down to Trump's favorite COVID-19 drugs
BY JOHN TRAVIS | JUN. 15, 2020

Could a global 'observatory' of blood help stop the next pandemic?
BY ROBERT BAZELL | JUN. 13, 2020

More ScienceInsider

Sifter

Three dozen alien civilizations may be advanced enough to communicate with us
BY DANIEL CLERY | JUN. 15, 2020

Rodent brains reveal triggers of hibernation
BY KELLY SERVICK | JUN. 12, 2020

Earth's species disappearing at an alarming rate
BY AMANDA HEIDT | JUN. 4, 2020

Watch an example of chimpanzee 'culture,' as one fishes for termites
BY AMANDA HEIDT | MAY 28, 2020

Our Moon is not as 'dry' as we thought
BY AMANDA HEIDT | MAY 8, 2020

More Sifter

LATEST CORONAVIRUS RESEARCH

Read the Latest Issue of *Science*

12 June 2020
Vol 368, Issue 6495

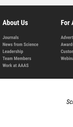


Table of Contents

MEDICINE/DISEASES
Computing cancer's weak spots

ASTRONOMY
Galactic flash points to source of enigmatic fast radio bursts

EPIDEMIOLOGY
Coronavirus rips through Dutch mink farms, triggering culls

ANTHROPOLOGY
Prominent Harvard anthropologist put on leave

MEDICINE/DISEASES
Vaccines that use human fetal cells draw fire

MEDICINE/DISEASES
Authors, elite journals under fire after major retractions

Get Our E-Alerts

Receive emails from *Science*. See full list
- Science Table of Contents
- Science Daily News
- Weekly News Roundup
- *Science* Editor's Choice
- First Release Notification
- Science Careers Job Seeker

United States
Email address*
I also wish to receive emails from AAAS/Science and *Science* advertisers, including information on products, services, and special offers which may include but are not limited to news, career information, & upcoming events.
Sign up today
Required fields are indicated by an asterisk (*)

Support nonprofit science journalism

Science's extensive COVID-19 coverage is free to all readers. To support our nonprofit science journalism, please **make a tax-deductible gift today.**

Donate | Not Now