

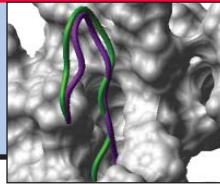
1502

The looming swine flu threat



1505

HIV's amazing escape act



1509

In the stem cell crossfire

HIV/AIDS

Vaccine Results Lose Significance Under Scrutiny

Everybody expected that the results of the first test of an AIDS vaccine would be controversial. Many researchers, after all, had been skeptical about the value of the vaccine for years and were waiting to pounce. But few would have predicted the confusion that ensued when VaxGen of Brisbane, California, released its long-awaited findings on 24 February (*Science*, 28 February, p. 1290). Over the next few days, VaxGen officials contradicted each other about how the data were analyzed and reported and whether some of the conclusions are statistically significant.

VaxGen's 5000-person study conclusively showed that the product failed to prevent infection with HIV, a result that sparked no disagreement. But company scientists claimed that analyses of subgroups revealed a statistically significant efficacy rate of 66.8% in blacks, Asians, and people of mixed race, and 78.3% in blacks alone. A few sharp-eyed scientists quickly questioned whether the company had calculated the statistics accurately.

Bette Korber, an HIV geneticist at Los Alamos National Laboratory in New Mexico, and biostatistician Steven Self of the University of Washington, Seattle, independently concluded that the subgroup analyses appeared to be flawed. Each time researchers conduct a sub-study of a data set, they increase the likelihood that an apparently significant result will be due to chance. To correct for this, statisticians assess a statistical penalty for each additional subgroup analysis. Specifically, statisticians adjust the confidence interval, which says with a given degree of certainty, or "*P* value," that the result is not due to chance. Both Korber and Self concluded that VaxGen had not adjusted for the subgroup analyses in the results the company reported.

On 26 February, Marc Gurwith, an infectious-disease specialist who heads the company's clinical trials, conceded to *Science* that because of a misunderstanding between the company's statistician and other scientists there, "the *P* values that were in

the [24 February] press release were not adjusted." Gurwith acknowledged that if a conservative correction were made, the significance of the result in blacks would disappear. Gurwith emphasizes, however, that statisticians do not have a single way to do these adjustments. "It's fairly complicated, because what the proper adjustments are is not so obvious," he says.

VaxGen originally claimed a *P* value of less than 0.02 for the black subgroup, meaning the company had a 98% or greater confidence that the result was not due to chance. Biologists typically use a *P* value of 0.05, or 95% confidence, as the dividing line between statistical significance and insignificance. Korber and Self argue that a widely used adjustment known as the Bonferroni correction should be applied to these results. The correction sim-

ply multiplies the *P* value by the number of subgroup analyses conducted.

sense," he says, noting that preliminary analyses show a correlation between anti-HIV antibody levels in vaccinated people and protection from HIV infection. Cornell University's John Moore, a longtime critic of the vaccine and an expert on HIV antibodies, finds this reasoning absurd. "Blacks and Asians lumped together is biological rubbish," says Moore. "They might as well do a subgroup analysis based on signs of the zodiac." And Self questions whether statistical significance holds up even in the combined group. "It's all murky because it's all post hoc analysis," says Self. "There's some marginal effect [of the vaccine], and it's worth going after, but it's not worth overblowing. It's a hypothesis-generating result."

The confusion over VaxGen's results took another odd turn on 27 February. That afternoon, VaxGen CEO Lance Gordon told an investor conference in New York City that not only had the company done a Bonferroni analysis, but "a conservative version was applied, and it had no impact on statistical significance. ... These are the accurate estimates, and the *P* values stand, even in view of subgroup and multiple subgroups." He asserted that the company had "proven" that its vaccine "does prevent HIV at least in those populations most responsive."

Later that day, VaxGen issued a terse press release, noting that the company had followed the analysis plan they had agreed on with the U.S. Food and Drug Administration. The number of necessary adjustments remains "subject to interpretation," the release says. "The company cannot predict the impact these adjustments may have on the findings, since that determination will ultimately rest with regulatory authorities."

Anthony Fauci, head of the National Institute of Allergy and Infectious Diseases, says he hopes the confusion will be cleared up at an AIDS meeting in Banff, Canada, that starts on 29 March: "Why don't we do what scientists always do, and settle this once and for all in an open forum?"

—JON COHEN



Close look. Claimed protection in blacks may not be statistically significant.

ply multiplies the *P* value by the number of subgroup analyses conducted.

Gurwith says VaxGen did nine substudies based on race. A Bonferroni correction would change the *P* value for the black subgroup to between 0.09 and 0.18. "So it wouldn't be significant," acknowledges Gurwith. He says the finding of significance in the group that combined blacks, Asians, and people of mixed race would remain, however. (Adjusted, *P* would be less than 0.04.) "This looks like a real result, and it makes some biological