STAT 100 Chance and Data Analysis October 4, 2002

Today: More on Design of Experiments:

1. Salk Polio Vaccine Experiment pp 3-142.
2. A Health Insurance Experiment pp 31-40

1. Salk Polio Vaccine Experiment pp 3-14

Problem: If polio occurs 1 per 5000 population per year approximately, and if I hope to prevent 50% of them, how many subjects do I need to demonstrate my prevention method convincingly in one year?

Let N be the population size. Then N x (1/5000) is the number of cases without vaccine
And N x (1/10000) is the number of cases with vaccine

If you chose N=20,000, the vaccine group might have about 2 cases, and the control group might have 4 cases for example. But this would not be convincing of the 50%.

Here are ten simulations of this situation, to see the variability that might be expected:

	Sim1	Sim2								Sim10
No V	1	7	3	3	6	5	4	1	9	3
Vacc	1	5	3	3	0	2	1	1	0	3

This simulation assumes the claim is correct, but any one simulation, based on testing two groups of 20,000 people, would not provide reliable evidence. But if we take the 10 simulations together, which is equivalent to using two groups of 200,000 people, we would find 42 cases in the control group (no V) and only 19 in the vaccinated group (Vacc). Repeating the above process for 200,000 people, 5 times, we get

	Sim1	Sim2			Sim 5
No V	41	36	40	33	36
Vacc	26	28	12	17	19

Now we see that the reduction in polio rate in the vaccine group is beginning to show through the random variation. If there were no reduction at all, we can simulate the situation where the polio rates in both groups is 1/5000, and typical results would look like this:

	Sim1	Sim2			Sim 5
No V	22	18	21	17	24
Vacc	19	23	26	16	20

The important point is that we can see that, IF the hoped-for reduction occurs, we would not get equivocal results like the ones in this last table. So if we see numbers like 42 cases in the No-vaccine group and only 19 in the Vaccine group, we would know that the vaccine was having at least some effect. But note that, even with 400,000 people in the study, we do not get a very good estimate of the percentage reduction - even though the long run reduction would be 50% by the design of our simulation, the rates based on 400,000 people varied in the five simulations from 22% (36 to 28) to 70% (40 to 12).

The need for 2,000,000 subjects should now seem reasonable.

Next Problem: If the polio rate varies as much as shown in Fig 1 (p 7), how is the effect of the vaccine going to be detected? A big change in the rate might occur even without the use of the vaccine!

This is why a **control group** is needed - it provides the background rate - at the same time and under the same conditions, except for the treatment of interest - against which the treatment can be compared.

Two **designs proposed**:

Observed-Control Approach: compare 1st and 3rd grade children (no vaccine) with 2nd grade children (vaccine)

Requires volunteers

Placebo- Control Approach: assign treatments (V or no-V) at random and use double-blind assessment.

Requires injection of placebo, more expensive design

BOTH designs were used. See Table 1 for results.

2. A Health Insurance Experiment pp 31-40

Question was: What is the effect on health and cost of the size of the deductible?

Design Issues: Sample Size: relevance of square root law. P 34
Allocation of treatments to subjects. Fewer to more expensive option.
Duration of Experiment: 3-5 years. Plan closure effect
Refusals: incentives needed.

Result: Less Cost, same health outcomes (Tables 1 and 2).

Methodological points: Use of simulation to assess sample size, placebo controls, sample size-precision link.