A Health Insurance Experiment

Joseph P. Newhouse *The RAND Corporation*

Most Americans have health insurance that helps them pay hospital and doctor bills. The key word is *helps*—only a very unusual insurance policy will pay the whole cost for all types of services. Most insurance policies require "cost sharing" by the patient through such mechanisms as deductibles and coinsurance. (A *deductible* of $200 per person per year requires the patient to pay the first $200 of medical bills in a year; *coinsurance* of 20% requires the person to pay 20% of every bill, above any deductible.) In some countries, however, public insurance may pay for all medical bills with no cost sharing.

Someone must decide how large deductibles and coinsurance rates will be— or whether there will be any cost sharing at all. The person deciding may be an employer or a union negotiator. In the Medicare and Medicaid programs, the decision makers are the Congress and the president. How should these people decide? What information should they want to know? It seems they should

want to know how much more health is bought by a more generous, and hence more expensive, insurance policy.

This issue has caused controversy for decades. Some have argued as follows: If people must pay for medical care, they will postpone or entirely forego needed services, and their health will be jeopardized. Those arguing this way usually conclude that medical care should be free. They tend to minimize the additional cost of care, taking the position that people do not enjoy going to the doctor and only go in times of real need.

Others have argued just the contrary: If medical care is free, people will abuse it and make many unnecessary doctor visits that do not affect their health. As a result, this group alleges that free medical care will be very costly, and they usually conclude that people should pay for most of their medical care, with insurance covering only very large medical bills.

Both sides have argued with more passion than evidence. To get more information about the issues, the federal government in 1974 set up a large experiment, the Health Insurance Experiment. People participating in this experiment were randomly assigned to different insurance plans that required them to pay differing amounts for their medical care. Both their health and the amount of medical care they used were monitored for several years. In its size and goals, this experiment in health policy was unprecedented.

Let's look at some examples of how statistics was used to design the experiment and to analyze the data collected in order to obtain as much information as possible for the money spent on the experiment.

THE STATE OF KNOWLEDGE AT THE BEGINNING OF THE EXPERIMENT

Why did we know so little in 1974? After all, health insurance was hardly novel. By 1974, when the experiment began, private insurance or the government had paid 80% to 90% of all hospital bills for two decades and insurance had paid over half of all doctor bills. One might have thought that the debate could be largely settled without an experiment simply by studying the actions of people with existing insurance policies.

Indeed, insurance companies could quote premiums on different insurance policies, and we could use those premiums to estimate the total cost to the person (including any cost sharing) of alternative insurance plans. But trying to shed light on cost and health effects of insurance by using data on existing insurance policies faces three difficulties:

1. People choose insurance policies with an eye to how much they would likely spend on medical care. Those who anticipate large expenditures have a clear incentive to seek out policies that cover their medical bills rather extensively. By contrast, those fortunate individuals who do not expect to see a doctor at all see little benefit in paying a large premium for nearly full coverage and so tend to opt for a policy with a small premium and large cost sharing. As

a result, if we try to infer the use of those not now on generous plans from the experience of those who have generous plans, or vice versa, we might considerably overstate the effects of insurance on use and cost.

2. Even if we could reliably assess the use of services (and thus the costs) under alternative insurance plans, there were no data on the effects of alternative plans on health. Yet this was the crux of the issue for many people who agreed that better insurance led people to seek help more often, but disagreed as to whether this help was necessary or unnecessary. Even if we obtained information on individuals' present health, insurance usually changes when people change jobs or retire; thus we could not infer whether present health was related to the present or the past insurance policy.

3. Finally, some insurance arrangements simply had not been tried. For example, some advocates of cost sharing thought cost sharing should be scaled down for the poor. But the poor typically either had no insurance or were covered in full by their state's Medicaid program. How the poor might respond to a plan that had them pay, say, up to the first 5% of their income for medical care could not be learned from existing data.

THE DESIGN OF A SOCIAL EXPERIMENT

Around 7,700 people who lived in six different areas of the country participated in the Health Insurance Experiment. They did not use their own health insurance (if any) for either three or five years and instead used an experimental health insurance plan. At the end of the three or five years, they returned to their usual health insurance.

Some participants received free medical care; others paid 25%, 50%, or 95% of their medical bills, up to a maximum of $1,000 out-of-pocket per family per year. This maximum was reduced for poorer families; specifically, it was never more than 5%, 10%, or 15% of a family's income, depending on the plan to which they were assigned.

This sketch of the Health Insurance Experiment has already encompassed a multitude of decisions on its design. Some of the issues decided include: (1) How many people will participate, and how many will be assigned to each plan? (2) How many sites will be used? (3) Exactly which people will be assigned to which plans? (4) How long will the experiment last? Statistical thinking helped the planners reach those decisions.

The Number of People in the Experiment

Because the objective of any experiment is to produce information, the experiment's designer seeks to maximize the information yielded for the monies spent and to inform those paying for it about how much more information they can buy for more money. One way to measure the amount of information is the precision of the results, or how likely it is that the results of the experiment (for example, in this experiment, the number of doctor visits per person per

year on an experimental plan) approximate the true state of the world (for example, how many times per person per year people would seek medical help if everyone had that insurance plan). The likelihood of the experimental results being close to the true number can always be increased by enrolling more people in the experiment. But we get diminishing returns from enrolling more people; for example, increasing the sample from 50 persons to 100 persons gains more than increasing it by another 50 persons to 150 persons. In fact, the precision of any given sample is generally proportional to the square root of the sample size.

Although larger samples give more precise results, they also cost more. Just as a person building a house must decide how much space to buy, the person paying for the experiment has to decide how much precision to buy; the experiment's designer cannot say how much precision should be bought, any more than an architect can say how large a house should be.

Sometimes one sees a statement that implies there is some critical threshold for sample size, and if the sample does not exceed that number, the study is not "scientifically valid." This notion is false; it is like saying a structure must exceed a certain number of square feet or it is not a house.

The Number of Persons per Plan

Now suppose our experiment's budget has been set. The designer must then decide how many people should be assigned to each experimental insurance plan. The first reaction might be to allocate equal numbers of people to each plan, but suppose one plan is much more costly per person enrolled. For example, the insurance plan with free medical care might be more costly than the plan that has persons pay 95% of the bill. Then we do not allocate equal numbers to plans but rather allocate them in accordance with the square root of the reciprocal of the cost ratio. For example, if one insurance plan were four times as costly as another, it would be assigned half as many people as the other plan. We use the square root of the cost ratio because the precision of a sample increases as the square root of the sample size.

The Number of Sites

Cost considerations also affect the appropriate number of sites. Certain costs must be paid just to enter a site; for example, a sampling plan must be drawn up and perhaps space for a site office must be rented. Hence, the more sites, the fewer total persons can be enrolled for any given budget. Our first thought might be to maximize the number of persons enrolled by choosing all persons to be in a single site, but we might have second thoughts when deciding which site to choose. If all sites were alike, one site would be optimal. But because sites differ (for example, in how long people stay in the hospital), we gain information by having more sites. In general, the more sites differ, the more sites and the fewer total people one should have. Cost is also relevant; the higher the costs of entering a new site relative to the costs of adding another person at an old site, the fewer sites are optimal.

The Length of the Experiment

The designer with a given budget must also decide whether to enroll fewer persons for a longer period of time or more persons for a shorter period of time. Two considerations favor enrolling more persons for a shorter period of time. A person's behavior tends to repeat itself over time (for example, someone with severe acne may make frequent doctor visits). Hence, two years of observation of one person is not as valuable as one year of observation of two persons because we already have a clue from the first year whether a person is a high user or not. Moreover, enrolling more persons for a shorter period of time makes information from the experiment available sooner.

But, pushed to its logical limit, these arguments imply enrolling as many persons as the budget would permit, each for one day! That seems absurd, and for good reason. Some of the behavior one wants to measure takes time to occur; for example, what happens to a person's health. Thus we want the experiment to be as short as possible, consistent with giving a long enough time period to allow the insurance plan to have its effect.

In choosing the period of time to operate the experiment, the designer must consider that persons enrolled in an experiment may not behave as they would if a program were of indefinite duration. For example, persons with an experimental insurance plan that will end on a certain date may crowd in or postpone certain services depending on what their situation will be after the experiment. Because this behavior would not be observed in an actual (nonexperimental) plan, it can contaminate the results. One strategy that was used in the Health Insurance Experiment to assess the amount of contamination was to enroll persons for different lengths of time, some for three years and some for five years. Any crowding in or postponement of services at the end of the experiment was measured by comparing the last year of a three-year group with the third year of the five-year group that began at the same time. Similarly, any initial surge was measured by enrolling in other sites a five-year group first and then a three-year group two years later.

The Refusals

The designer of any experiment must take into account that (generally) people cannot be forced to participate in an experiment but must instead agree to participate. Moreover, the designer is under an ethical and legal injunction to inform potential participants about the benefits and risks of participating. Although participants must agree and must be informed, the designer controls the benefits, and perhaps the risks, which will influence the potential participants' decisions.

To minimize refusals, the Health Insurance Experiment paid people enough to ensure that they could not lose financially by participating. For example, on one experimental plan the person had to pay 50% of any hospital bill up to a $1,000 maximum in any year. Suppose the person's prior insurance plan would have fully covered any hospital stay. In this case the experiment paid the person $1,000 per year, irrespective of whether the person went to the

hospital. If the amount paid had been less than $1,000, persons who expected to go to the hospital with a high degree of certainty (for example, pregnant women) would probably refuse to enroll because they would expect to lose money.

To be sure, the separate payment might itself change behavior—and, if it did, it would be important to know about it because the separate payment would not be part of a "real world" insurance plan. In order to study the payment's effect, the designer can build some variation into the payment, for example, by randomly choosing some families to receive $1,500 rather than $1,000. Such variation was in fact built into the Health Insurance Experiment, and it turned out that the payments had a negligible effect on behavior.

THE ANALYSIS OF THE EXPERIMENTAL DATA

Table 1 shows the use and spending on medical care per person by insurance plan in the Health Insurance Experiment. Clearly how well a person is insured matters a great deal in how many medical care services he or she uses. Persons paying 95% of the bill made 40% fewer doctor visits than those paying nothing. Similarly, hospital usage decreased by 23%, and total spending by 31%.

We want to assess whether the additional services received by the group on the free care plan affected their health. Reliable evidence on this issue had never before been available. The data in Table 2 show that the health of the average person was little affected by the differences in use highlighted in Table 1. (Table 2 groups together the 25%, 50%, and 95% coinsurance plans for ease of interpretation. The outcomes of these three plans were similar on all measures.)

The values shown in Table 2 are three of the many measures that were collected; they include the two measures, blood pressure and vision, where we can be reasonably confident that additional services did lead to an improvement in health. (Certain measures of dental health were also improved.) The overall measure of health, General Health Perceptions, is, however, slightly better

Table 1   Measures of medical care use, per person per year, as a function of insurance*

Percentage of Bill Paid by Person†	Physician Visits	Hospital Admissions per 100 Persons	Spending‡ (1984 Dollars)
0	4.55	12.8	777
25	3.33	10.5	630
50	3.03	9.2	583
95	2.73	9.9	534

*Although not shown, the precision of these numbers is quite good. For details, see Manning et al. (1987).
†Percentage was paid up to an annual out-of-pocket maximum, which was $1,000, or 5%, 10%, or 15% of income, whichever was less. The percentage of income was assigned at random.
‡Includes monies paid by insurance and by the person. Excludes spending on dental services and outpatient psychotherapy. Dental services and outpatient psychotherapy exhibit about the same degree of responsiveness to plan; including them would add a little more than one-third to each of the numbers. Most of the additional monies would be for dental services. The numbers in this column come from averaging predicted spending; see Manning et al. (1987).

Table 2   Measures of health outcome as a function of insurance plan*

Type of Insurance Plan	General Health Perceptions†	Diastolic Blood Pressure (mm of Hg)	Vision Measure‡ (Snellen Lines)
Free Care	67.4	77.9	2.42
Cost Sharing	68.0	78.8	2.52

*These values are predicted from a regression equation and hence are somewhat more precise than simple means. For details, see Brook et al. (1983).
†This is a scale that ranges from 0 to 100 (100 being healthiest) and that describes how people rated their own health. It is based on answers to 22 different questions. A 5-point difference in the scale, all else equal, is equivalent to having mild hypertension (high blood pressure); a 7-point difference, all else equal, is equivalent to having diabetes mellitus.
‡An outcome of 2 corresponds to 20/20 corrected vision in the better eye, an outcome of 3 corresponds to 20/25, and an outcome of 4 to 20/30; thus, smaller numbers are preferable.

for the group of individuals who paid something for their medical care (68.0 versus 67.4). More important, though, is the question of how precise this estimate is. Using statistical analysis we can conclude that any true difference in General Health Perceptions between the two groups is likely to be small; specifically, it is likely to be between 1.5 points (favoring cost sharing) and −0.3 point (favoring free care). More specifically, we can say that if a similar experiment was carried out many times, and the same procedure was used to construct this band, 95% of the resulting bands (there may, of course, be no difference). Because the upper limit of the band in this case is 0.3 point, we can be reasonably certain that the free plan is unlikely to be more than 0.3 point better than the cost-sharing plan.

Of course, we have no intuition about what a 0.3-point difference might mean. One way to gain some intuition is to look at all the people in the study who have some disease, for example, hypertension (high blood pressure). If we adjust for observable differences between people who do and do not have hypertension (that is, adjust for the older age of the group with hypertension), the group with hypertension has a value for General Health Perceptions about 5 points less than the value for the group without hypertension. (For an explanation of adjustment see the Moses and Mosteller essay.) Because 5 points is large relative to 0.3 point, one interpretation of these outcomes is that the additional services that the people on the free care plan consumed did not make them noticeably healthier in their eyes. They may have thought themselves a tiny bit healthier and the measurements failed to detect that, but the opposite may also be true.

For people with hypertension and vision problems, we can be relatively sure that beneficial effects occurred. Those effects were concentrated among the group of poor individuals.

Upon reflection it is probably not surprising that these two conditions benefited from additional care. Hypertension and vision problems (and dental disease as well) are widespread relative to other diseases and are relatively inexpensive to diagnose and treat. Moreover, it is plausible that the poor would benefit differentially. A thoughtful reader might ask why, if these effects on

blood pressure and vision existed, they were not reflected in a measure such as the General Health Perceptions. One answer is that the proportion of people affected by these problems and the benefits of free care over those of cost sharing for those problems are sufficiently small that, when a measure of health is averaged over the entire population, the effects are not detectable with the sample size in the experiment. Another possibility is that some fraction of the additional services on the free care plan were associated with bad outcomes (for example, reactions to a drug that was prescribed). These bad outcomes may have offset other good outcomes.

THE INFLUENCE OF THE RESULTS

Because the experiment's findings were important and unique, especially those in Table 2, they appear to have had a substantial effect. The results in Table 1 were published in December 1981, and those in Table 2 in December 1983. Between 1982 and 1984 there was a considerable increase in the amount of cost sharing in private health insurance in the United States, especially for hospital services. Rates of hospital admissions declined, as did physician visits. Table 3 shows these changes. Using the results of the experiment, we can infer that reductions in use in response to greater cost sharing in private insurance probably had little or no adverse effect on health—because most people with private insurance are not poor.

Of course, we cannot attribute all the changes shown in Table 3 to the information provided by this experiment; indeed, we can never know how much of a role the experiment played in bringing them about. But the experiment's results are well known to consultants who advise companies on their health

Table 3   Change between 1982 and 1984 in insurance and medical care utilization

	1982	1984
Percent of private insurance plans with a deductible for hospital services	30	63
Percent of private insurance plans with a deductible of $200 per person per year or more	4	21
Hospital discharges per 100 persons under 65*	11.8	10.7
Physician visit rate per person under 65	4.9	4.7

*Excludes deliveries.

Source: First two rows: Hewitt Associates, cited in Jeff Goldsmith, "Death of a Paradigm: The Challenge of Competition," Health Affairs, 3(3:Fall 1984):p.12. Second two rows: National Center for Health Statistics, National Health Survey, various issues.

plan benefits, and some companies even cited the results in brochures to their employees explaining the changes. Thus it is plausible that the experiment played an important role in bringing about these changes.

The experiment, all told, cost around $80 million. Was the information worth that kind of money? Although $80 million is certainly a substantial sum, it pales by comparison with the $230 billion that the United States spent on hospital and physician services in 1984 alone. Table 3 shows that hospital discharges fell about 10% and physician visits fell about 4% between 1982 and 1984. If the results of the experiment caused even a small portion of this change—and few or no effects on health—then the experiment would have paid for itself in the space of a few months!

PROBLEMS

1. The following debate has occurred over the pricing of telephone service: Some people think that with a flat rate per month for local service, people spend much more time on the telephone than they would if they paid per minute. Others say the effect isn't very great. Do you think an experiment would be a good way to assess how people's use of the telephone might change if they had to pay by the minute? Why or why not? (Consider that some households now pay by the minute while other households pay a flat rate.)

2. Suppose you were going to conduct an experiment to assess how much telephone use responded to ending flat rate charges. Name some design decisions you would have to make. How would you induce people to participate?

3. Do you think an experiment would be the best way to assess the effect of the spread of health maintenance organizations on medical care prices and costs? Why or why not?

4. How does precision change with sample size? Is there a sample size that makes a study scientifically valid? Explain your answer.

5. When should the same number of people be assigned to each experimental treatment?

6. How should the number of sites in an experiment be chosen?

7. How long should an experiment run? Explain your answer.

8. What are the threats to the validity of experimental results from refusals to participate in an experiment?

REFERENCES

R. Archibald and J. P. Newhouse. In press. "Social Experimentation: Some Why's and How's." In E. Quade and H. Miser, eds. Handbook of Systems Analysis, Vol. II. New York: North Holland. (Also available as Rand publication R-2479, May 1980. Santa Monica, CA: Rand Corporation.)

R. H. Brook, J. E. Ware, Jr., W. H. Rogers, E. B. Keeler, A. R. Davies, C. A. Donald, G. A. Goldberg, K. N. Lohr, P. C. Masthay, and J. P. Newhouse. 1983. "Does Free Care Improve Adults' Health? Results from a Randomized Controlled Trial." *New England Journal of Medicine* 309:1,426–1,434.

W. G. Manning, J. P. Newhouse, N. Duan, E. B. Keeler, A. Leibowitz, and M. S. Marquis. 1987. "Health Insurance and the Demand for Medical Care: Results from a Randomized Experiment." *American Economic Review* 77:251–277.