SAFETY OF ANESTHETICS

Lincoln E. Moses *Stanford University*
Frederick Mosteller *Harvard University*

In 1958, American hospitals began using a new anesthetic called halothane. It soon became widely accepted for its many desirable properties. Unlike some of the more commonly used anesthetics, it could not catch fire, so fire and explosive hazards did not have to be a concern during surgical operations. Patients found it less disagreeable and recovered from anesthesia more quickly and with less severe aftereffects. Extensive laboratory research and trials on animals and humans in surgery had encouraged belief in its safety. So there were good reasons for halothane to come rapidly into widespread use. By 1962, surgeons used halothane in half of their operations.

After a few years, however, halothane came under suspicion as accounts appeared in the medical literature of some strikingly unusual—but strikingly similar—deaths of patients who had recently had this anesthetic. A few patients recovering from surgery suddenly took a turn for the worse, ran fevers, and died; subsequent autopsies revealed massive fatal changes in their livers. Even though there were only a few such reports, it was natural to ask, "Do these incidents mean that halothane is a poison dangerous to people's livers? Should the use of halothane as an anesthetic in surgery therefore be discontinued?" Some compounds chemically similar to halothane were already known to cause liver damage, so the question arose with all the more force. A national committee was appointed to assemble and examine evidence that might help to answer these questions.

Much information was needed. First, it was possible that the liver changes found with halothane might occur equally often with other anesthetics, but that physicians rarely reported such cases because the old anesthetics would attract less attention. Second, and this is a key point, whether or not halothane had a special adverse effect on the liver, there was the possibility that its other advantages might result in a lower overall postoperative death rate. Frequently things that are easy to use actually work better (this is true of sharp knives, fine violins, and easy-to-read instructions), so it might well be that an anesthetic that is easier for the doctor to use might work better and result in somewhat lower death rates during surgery. Information was needed to confirm or refute this possibility.

These questions could not be answered by doing laboratory experiments with mice, for mice might well behave differently under the anesthetic than human beings; nor could the questions be answered by looking in books, for the information was not even known; nor was it useful to ask experts for their judgments, since different experts had widely different opinions.

THE EVIDENCE

It was necessary to amass a great deal of evidence, so the committee decided to conduct a survey of hospital experience. At that time, halothane had been given to about 10 million people in the U. S. Many of these patients had been in a particular group of 34 hospitals that kept good records and whose staffs keenly wanted to answer the very questions we have asked. They cooperated with the committee by sampling their records of surgical operations performed during the years 1960–64. In addition to recording whether or not the patient died within six weeks of surgery, they gave information on the anesthetic that had been used as well as facts about the surgical procedure and the patient's sex, age, and physical status prior to the operation. Among the 850,000 operations in the study, there were about 17,000 deaths, or a death rate of 2%. The death rates, shown in Table 1, were calculated for each of four major anesthetics and for a fifth group consisting of all other anesthetics. Note that these are death rates from all causes, including the patients' diseases, and are not deaths especially resulting from the anesthetic.

TABLE 1. Death Rates Associated with Various Anesthetics

HALOTHANE	PENTOTHAL*	CYCLOPROPANE	ETHER	ALL OTHER
1.7%	1.7%	3.4%	1.9%	3.0%

* Nitrous oxide plus barbiturate.

NEED FOR ADJUSTMENT

Table 1 suggests that halothane was as safe as any other anesthetic in wide use, but such a suggestion simply cannot be trusted. Medical people know that certain anesthetics, cyclopropane, for example, are used more often in severe and risky operations than some other anesthetics (such as pentothal, which is much less often used in difficult cases). So some or all of the differences between these death rates might be due to a tendency to use one anesthetic in difficult operations and another in easier ones.

Different operations carry very different risks of death. Indeed, death rates on some operations were found to be as low as 0.25% and on others nearly 14%. The change in death rates across the categories of patients' physical status was even more dramatic, ranging from 0.25% in the most favorable physical status category to over 30% in the least favorable. Age mattered a great deal: the most favorable 10-year age group (10 to 19) carried a death rate of less than 0.50%, while the least favorable age group (over 90) had a death rate of 26%. Another factor was sex, with women about two-thirds as likely to die. Because the factors of age, type of operation, physical status, and sex were so very important in determining the death rate, it was clear that even a relatively small preponderance of unfavorable patients in the group receiving a particular anesthetic could raise its death rate quite substantially. Thus, the differences in Table 1 *might* be due largely, and possibly even entirely, to discrepancies in the kinds of patients and types of operations associated with the various anesthetics. Certainly it was necessary to somehow adjust, to equalize for, type of operation, sex, physical status, and age before trying to determine the relative safety of the anesthetics.

If these data had been obtained from a properly planned experiment, the investigators might have arranged to collect the data so that patients receiving the various anesthetics were comparable in age, type of operation, sex, and physical status. As it was, these data were collected by reviewing old records, so substantial differences in the kinds of patients receiving the various anesthetics were to be expected and, indeed, were found. For example, cyclopropane was given two or three times more often than halothane to patients with bad physical status and substantially more often than halothane to patients over sixty years of age. There were many such peculiarities, and they were bound to affect the death rates of Table 1.

CARRYING OUT ADJUSTMENTS

The task was to purge the effects of these interfering variables from the death rates corresponding to the five anesthetics. Fortunately, it was possible to do this. We said "fortunately" because it could have been impossible; for example, if each anesthetic had been applied to a special set of operations, different from those in which other anesthetics were used, then differences found in the death rates could perfectly well have come from differences in the operations performed and there would be no way to disentangle operation from anesthetic and settle the question. But in the data of this study, there was much overlapping among the anesthetics in the categories of age, kind of operation, sex, and patient's physical status, so that statistical adjustment, or equalization, was possible. Analysis using such adjustment was undertaken in a variety of ways because the complexity of the problem made several different approaches reasonable and no one approach alone could be relied upon. There was close agreement in the findings regardless of the method used.

A summary of the results is shown in Table 2. Notice that halothane and pentothal after adjustment have higher death rates than their unadjusted rates in Table 1. Also, cyclopropane and "all other," the two with highest rates in Table 1, now have lower death rates. The effects of these adjustments are quite important: halothane, instead of appearing to be twice as safe as cyclopropane—the message in Table 1—now appears to be safer by only about one-fifth.

Figure 1 shows the adjustment effect graphically.

A Special, Simple Case. To understand the nature of the adjustments, it is helpful to look at a very special case with fictitious data.[1] Suppose that halothane and cyclopropane are to be compared, that we are to adjust only for physical status, and that we classify all patients in either good or poor physical

[1] This section may be skipped without disturbing continuity.

TABLE 2. Adjusted Death Rates for Various Anesthetics (Adjusted for Age, Type of Operation, Sex, and Physical Status)

HALOTHANE	PENTOTHAL	CYCLOPROPANE	ETHER	ALL OTHER
2.1%	2.0%	2.6%	2.0%	2.5%

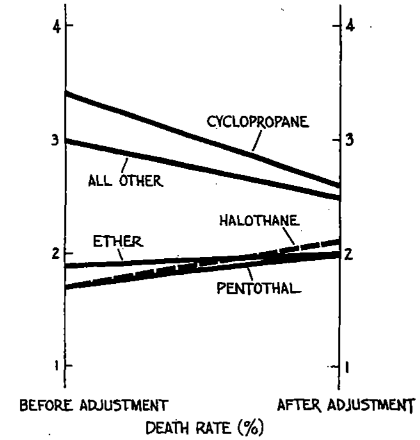

FIGURE 1
Effects of adjusting death rates. Plotted points are from Tables 1 and 2

status. The fictitious data might look like those shown in Table 3. Thus, halothane was given to patients in good status four out of five times (in this fictitious case), but cyclopropane, just the other way around, was given to patients in good physical status one out of five times. The two anesthetics have exactly the same death rates for each physical status classification separately, but the very different proportions of status for the two make the overall death rate for cyclopropane almost twice that of halothane.

Suppose that we decide to adjust by computing what the overall death rates *would* be if the two anesthetics were given to a population of patients in which 70% were of good status and 30% poor. The death rates for each status group remain the same. The computations would be as displayed in Table 4. For each anesthetic the overall death rate is now 1.6% with these fictitious data.

TABLE 3. Fictitious Data

	NUMBER OF PATIENTS	NUMBER OF DEATHS	DEATH RATE
Halothane			
Poor physical status	200	6	3.0%
Good physical status	800	8	1.0%
All patients	1000	14	1.4%
Cyclopropane			
Poor physical status	800	24	3.0%
Good physical status	200	2	1.0%
All patients	1000	26	2.6%

TABLE 4. Adjusted Death Rates (Fictitious Data)

	NUMBER OF PATIENTS	NUMBER OF DEATHS	DEATH RATE
Halothane			
Poor physical status	300	9	3.0%
Good physical status	700	7	1.0%
All patients	1000	16	1.6%
Cyclopropane			
Poor physical status	300	9	3.0%
Good physical status	700	7	1.0%
All patients	1000	16	1.6%

In practice, the actual adjustments are far more complicated, partly because fine breakdowns of the data leave many cells with no entries at all and many with small entries.

Consistency over Hospitals. Even after the adjustments, there was still considerable variability from hospital to hospital. Our confidence in the validity of the statistically adjusted figures in Table 2 is affected by how consistent the death rates were from hospital to hospital. After all, if there were many hospitals where cyclopropane had a lower adjusted death rate than did halothane, even though halothane was lower "on the average," we might feel uncertain that halothane was really safer just because of the difference in adjusted death rates. Some fluctuation from hospital to hospital is to be expected, of course, because of chance factors; absolute consistency is not to be expected.

The question is: "Were the comparisons between the adjusted anesthetic death rates consistent enough over hospitals to warrant taking seriously the apparent differences shown in Table 2?" Rather complicated statistical techniques were necessary to study this question, but the conclusions were clear: halothane and pentothal both had adjusted death rates definitely lower than the adjusted death rates for cyclopropane and "all other," and those lower adjusted death rates were real in the sense that they could not be explained by chance fluctuations. The differences were sufficiently consistent to be believed. On the other hand, ether, with the same adjusted death rate as halothane and pentothal, did not have a consistent pattern of comparison. Therefore, ether cannot be reliably compared with the others; we cannot tell from the evidence obtained whether it may be somewhat safer than it appears here—or somewhat less safe. Possibly it is as safe as, or safer than, halothane; possibly it is no safer than "all other." The findings about ether are indefinite because there were fewer administrations of ether and these were concentrated in a few hospitals; hence there are fewer data to go on.

A SUMMING UP SO FAR

What can we conclude so far from the findings of this study? First, and most important, is the surprising result that halothane, which was suspect at the beginning of the study, emerged as a definitely safe and probably superior anesthetic agent. Second, we see that a careful statistical study of 850,000 operations enabled the medical profession to answer questions more firmly than had previously been possible, despite the much greater "experience" of 10 million administrations of halothane and other 10s of millions of administrations of the other anesthetics.

HOSPITAL DIFFERENCES AGAIN

A brand new question emerged with the observation that the 34 hospitals had very different overall postoperative death rates! These ranged from around 0.25% to around 6.5%. This seemed to mean that the likelihood of dying within six weeks after surgery could be more than 20 times as great in one hospital as in another. Just as before, however, there were strong reasons to approach this startling information skeptically. Some of the hospitals in the study did not undertake difficult operations, such as open-heart surgery, while others had quite large loads in such categories. This kind of difference alone would cause differences in hospital death rates. Further, the age distribution might be different, and perhaps importantly so, from one hospital to another. Indeed, one was a children's hospital, another a veteran's hospital. Some hospitals might more frequently accept surgical patients having poor physical status. So the same interfering variables as before surely affected the differences in hospital death rates. If adjustment was made for them, would the great differences in hospital death rates vanish? Be much reduced? Remain the same? Or, as is conceivable, actually increase?

Adjustment procedures were applied, and the result was that the high-death-rate hospitals, after adjustment, moved down toward a 2% overall death rate, and the low-death-rate hospitals, after adjustment, moved up toward a 2% overall death rate. The adjusted hospital death rates no longer ranged from 0.25% to over 6.5%; instead, the largest of these adjusted death rates was only about three times as great as the smallest. Now, almost any group of 34 such rates will exhibit some variability, and the ratio of the largest to the smallest must be a number larger than 1. Even if a single hospital were measured over several different periods, the rates would fluctuate from chance alone. The fluctuation in rate would be considerable because the death rate itself is basically low and one death more or less makes a difference in the observed rate for the period. The fact that this ratio turned out to be 3 in these data does not, in itself, indicate clearly that there were real, unexplained hospital differences.

Careful statistical study showed that there are probably some real differences from hospital to hospital in postoperative death rates, and that these differences cannot be explained wholly by the hospitals' patient populations in terms of age, sex, physical status, and surgical procedure. Statistical theory showed that the ratio of highest to lowest adjusted rate should be about 1.5 if the hospitals were identical in operative death rate after adjustment (so that the adjusted rates differ only by chance fluctuation).

The position, then, is that we began with the large ratio of about 25 for unadjusted rates, cut the ratio way down to 3 for adjusted rates, and then compared the 3 with 1.5 as a theoretical ratio. Because the adjustments can hardly have been perfect and because there undoubtedly were unadjusted factors that differed among the hospitals, we conclude that the adjusted hospital death rates are indeed close together. Thus, what, on the basis of the unadjusted hospital death rates, looked like a shocking public health problem proved, after statistical investigation, to be quite something else. The apparent problem was mainly, though perhaps not entirely, a dramatic manifestation of differences not in the quality of surgical care, but in the difficulty of the surgical cases handled in the various hospitals.

LIVER DAMAGE

Let us return to the question of liver damage. The total deaths from this source were few, and the lack of autopsies for nearly half of the deaths made firm conclusions impossible. Since the study was made, however, an anesthesiologist who had often been exposed to halothane while administering it to patients, has been discovered to be sensitive to halothane; he exhibits symptoms of liver malfunction from breathing it.

Are occasional individuals sensitive to other anesthetics? How many people develop such sensitivity? These are hard questions to answer, and they are especially difficult to study because of the rarity of the occurrences.

CONTRIBUTIONS OF STATISTICS

What were the main contributions of statistics in the program? First was the basic concept of a death-rate study. This needed to be carried out so that the safety of anesthetics could be seen in light of total surgical experience, not just in deaths from a single rare cause. Second, though we have not discussed it, the study used a special statistical technique of sampling records designed to save money and to produce a high quality of information. Third, special statistical adjustments had to be created to appraise the results when so many important variables—age, type of operation, sex, patient's physical status, and so on—were uncontrolled. Fourth, as a result, the original premise of the study was not sustained, but a new result emerged: halothane seemed safer than cyclopropane. The study produced a nonfact as well: the comparative merit of ether is uncertain. We cannot tell if its associated death rate is higher, lower, or nearly the same as halothane, though the indication is that it is nearly the same. Fifth, new evidence on hospital differences showed that the initial, wide variation in death rates, when suitably adjusted for patient populations, left only small unexplained differences in rates. These remaining differences led the medical profession to begin, in 1971, a study of possible causes of hospital-to-hospital differences in postoperative death rates, in the hope of discovering ways to improve postoperative care.

PROBLEM

1. Explain how it was possible for halothane to be as safe as other anesthetics despite the evidence from autopsies that it caused liver damage.
2. What is meant by equalization or adjustment?
3. Explain the discrepancy between the safety of halothane indicated in Tables 1 and 2.
4. How can you explain the unadjusted differences of the overall postoperative death rates in the 34 hospitals in the study?
5. What were the principal conclusions of the study?
6. What statistical tools were used in the study?
7. Explain why the results about ether were uncertain.
8. Is halothane a cause of liver damage? Explain.
9. Refer to Figure 1. For which anesthetic is the adjustment most drastic? Check your answer by comparing Table 2 with Table 1.
10. In Table 4 explain how the number of deaths in each status group was calculated.
11. Consider a study of the effect of a high cholesterol diet on mortality from coronary heart disease. The mortality rates from CHD in two groups are compared—one group with high cholesterol diet and another with an average diet. What are some of the factors that one needs to adjust for in the two groups before comparing the mortality rates?

12. Why did the statistical study described in this article concentrate on 850,000 operations only although there were over 10 million operations in which halothane was administered and 10's of millions of operations where other anesthetics were used?

REFERENCE

John P. Bunker, William H. Forrest, Jr., Frederick Mosteller, and Leroy D. Vandam. 1969. *The National Halothane Study,* Report of the Subcommittee on the National Halothane Study of the Committee on Anesthesia, Division of Medical Sciences, National Academy of Sciences—National Research Council, published by the U. S. Government Printing Office, Washington, D. C., for the National Institutes of Health, National Institute of General Medical Sciences, Bethesda, Md.