If Nothing Goes Wrong, Is Everything All Right?

Interpreting Zero Numerators

James A. Hanley, PhD, Abby Lippman-Hand, PhD

PHYSICIANS frequently must provide patients with estimates of the risk of a particular medical procedure or of the probability of a specific health outcome. To do so, they may use data from their own experiences as well as those available in the literature. A general goal is to locate the estimate within a fairly narrow range.

When the procedure or outcome of concern is relatively uncommon, precise estimation, although desirable, may be difficult or impossible. Moreover, when the only data available come from a study in which none of the events of concern actually occurred—ie, a study in which a zero numerator is reported—making inferences seems to be particularly problematic. Because the occurrence of "no events" seems to be viewed as very different both quantitatively and qualitatively from the occurrence of one or more events, it is useful to look into some of the statistical and psychological issues that influence the interpretation of a zero numerator. In particular, we would like to emphasize that (1) a zero numerator does not necessarily mean "no risk," (2) a zero numerator does not preclude inferences about the size of a risk, and (3) the principles of inferential statistics that apply to nonzero numerators apply equally well to zero numerators. In fact, there is a quick and simple rule that establishes the maximum long-run risk associated with an observation of no effects in a sample of any given size.

From the Department of Epidemiology and Health (Drs Hanley and Lippman-Hand) and the Centre for Human Genetics (Dr Lippman-Hand), McGill University, Montreal.

Reprint requests to Department of Epidemiology and Health, McGill University, 3775 University St, Montreal, Quebec, Canada H3A 2B4 (Dr Hanley).

Examples

Studies reporting a zero numerator are fairly common in the literature. The following examples are from recent issues of major medical journals, including this one. They have been arbitrarily chosen strictly for illustrative purposes. Thus, we have reduced the details to a minimum. For each, the reader should consider the inference that might be made from these data.

EXAMPLE 1.—Of 14 boys followed up for a median of 5½ years after chemotherapy for leukemia, none had abnormal testicular function (ie, the abnormality rate was 0/14).[1] With what risk, if any, of testicular dysfunction might these results be compatible?

EXAMPLE 2.—The status of 112 liveborn children whose mothers had been immunized against rubella was studied to assess the risks of gestational exposure to the vaccine.[2] None of the infants born (0/112) had any congenital malformations associated with congenital rubella. What is the maximum malformation risk compatible with finding none of 112 infants with defects in a single study?

EXAMPLE 3.—The final example is one we will adapt to explain the inferences that may be made when a zero numerator is found. In a study of siblings of 167 infants with tracheoesophageal dysraphism (TED), none was found to have a neural tube defect.[3] (We ignore here problems in defining the denominator for this study and assume, for illustrative purposes, that 167 siblings were studied and that the observed rate of neural tube defect was 0/167.) Is this evidence sufficient to say that the risk of neural tube defect in siblings of children with TED is not increased over the risk in the general population (1.5 per 1,000 births in the area studied)? Or is a 0/167 rate also compatible with a risk that would make parents of children with TED eligible for the prenatal diagnosis of neural tube defects in subsequent pregnancies?

Despite differences in focus, the three examples have the same structure: in each study, no adverse effects were found—nothing happened. Only the denominators—the sample sizes—differ. To illustrate the inferential process employed when there is a zero numerator, we will use the numbers in the third example and consider how these data might be used to evaluate the risk associated

with a contrast medium to be used by radiologists. (We chose a hypothetical example to avoid controversy about the specific details of any reported study.)

Suppose that the standard contrast agent used by radiologists over a long period has been shown to cause a serious reaction in about 15 of every 10,000 patients exposed to it. That is, the known risk with the old agent is 1.5 per 1,000 (the same as the population risk for neural tube defects in the third study cited). Suppose further that a new contrast agent is introduced. Soon afterward, a report of its use in 167 patients appears: no patient has had the reaction of concern. Can we infer from these data that, in the long run, the new medium will carry less risk, more risk, or the same risk as the old agent?

We can begin by assuming that the long-run risk with the new agent is no different from that with the old. Under this trial assumption, each patient has a 0.0015 chance of reacting and, conversely, a 1−0.0015= 0.9985 chance of not reacting. The chance that all 167 patients will avoid a reaction—that the reaction rate would be 0/167—is 0.9985^{167}, or 78%. Thus, finding no reactions among this group is not at all surprising. In fact, if the new agent carries the same risk as the old, it would be more surprising if anyone *did* react.

If we are more pessimistic and believe that the long-run complication rate with this new agent could be much higher than that with the old, we can assess a different trial value, for example, a risk of one in 100. Here again, however, observing no reactions in 167 patients would still not be *that* surprising: there would be a 0.99^{167}, or 19%, chance that none of these 167 patients would react to the new agent even if the true risk were one in 100.

Finally, we can take an even greater trial value, say a risk of four in 100. Were this the true risk, then there would be just a 0.1% chance of finding no reactions in a group of 167 patients, and we would express great surprise at an observation as discrepant with the "truth" as this.

How high we would go with these trial estimates of the true long-run risk depends on what might be viewed as our "willingness to be surprised." This is the point at which we decide that because a finding that should not happen (based on what we posit as the truth) seems nevertheless to have happened, it is not just because we are (un)lucky. Rather, since the observation really is too surprising if the long-run risk we posited were correct, we decide that the real risk must be lower.

As a matter of convention, a chance of 5% has come to be used as a limit to credibility; thus, a finding that had at least a 5% chance of occurring is considered not that surprising ("it can happen"), but anything more extreme is. Table 1 presents a range of trial values for the long-run risk and the chance of observing zero complications in 167 patients associated with each. As Table 1 shows, our limit for surprise of 5% is reached when the true long-run risk is about two in 100. In other words, the findings "fit" or are "not surprisingly different from" any long-run risk of 2/100 or less. This "plausible range of possibilities for the truth," which extends from zero to the maximum compatible with our 5% level of surprise, is usually termed a "95% confidence interval."

Table 1.—Probability That a Sample of 167 Will Have No Complications

If Long-Run Complication Rate Is	Then Probability (%) of Observing Zero Complications in 167 Is
1 in 10,000	98
1 in 1,000	85
1.5 in 1,000	78
1 in 200	43
1 in 100	19
1 in 56	5
1 in 25	0.1

Table 2.—Long-Run Risk Ruled Out by $0/n$ (With 95% Confidence, ie, 5% Limit of Credibility)

Rate of	Rules Out Any Long-Run Rate (%) Higher Than
0/10	26*(30)†
0/20	14(15)
0/30	10(10)
0/50	6(6)
0/100	3(3)
0/1,000	0.3(0.3)

*Derived "exactly" by solving (1−maximum rate)^n=0.05.
†Derived from rule of three.

Making Inferences: The Rule of Three

In all of these examples, the event of interest failed to occur in a finite number of subjects. In making inferences from observations with zero numerators, we want to say with a certain degree of confidence that the true or long-run risk (of chemotherapy, of rubella immunization, of having a sibling with NTD) is between zero and some upper limit.

To be 95% confident that our interval estimate of the long-run risk is correct, a simple rule (of unknown origin) can be applied.[4] This "rule of three" states that if none of n patients shows the event about which we are concerned, we can be 95% confident that the chance of this event is at most three in n (ie, $3/n$). In other words, the upper 95% confidence limit of a $0/n$ rate is approximately $3/n$. (This approximation is remarkably good: when n is larger than 30, the rule of three agrees with the exact calculation to the nearest percentage point; below 30, it slightly overestimates the risk, but then the maximum risk [ie, 26% if n=10] is becoming so high that the difference between it and that suggested by the rule [3/10=30%] is not usually worth quibbling about.) The derivation of the "rule of three" is as follows:

To find the largest risk with which the finding of $0/n$ is compatible (ie, a level of credibility of at least 5% or 0.05), one must either proceed by trial and error, as we did, or else explicitly solve the following equation:

$$(1-\text{Maximum Risk})^n=0.05$$

This can be rewritten, by taking the nth root of each side, as

$$1-\text{Maximum Risk}=\sqrt[n]{0.05} \qquad (1)$$

The reader who uses a calculator to evaluate $\sqrt[n]{0.05}$ or $0.05^{1/n}$ for n above 30 will find that it comes remarkably close to $(n-3)/n$ or $1-(3/n)$; thus, equation 1 can be rewritten as follows:

$$1-\text{Maximum Risk}=1-(3/n)$$

or

$$\text{Maximum Risk}=3/n$$

The reader who would like to understand "how the 3 got in there" can do so by writing $0.05^{1/n}$ as an infinite series, ie,

$$0.05^{1/n}=1+[(\ln 0.05)/n]+[(\ln 0.05)^2/2n^2]+\ldots \qquad (2)$$

where ln 0.05 refers to the natural logarithm of 0.05. Given that ln 0.05 is −2.9957, or −3 when rounded off to two decimal places, and given an n of 30 or more, the terms involving divisors of n^2 or bigger in the right-hand side of equation 2 make almost no contribution (ie, $0.05^{1/n}$ is very close to $1-[3/n]$).

For a 99% confidence interval, the cor-

responding shortcut is a "rule of 4.6" (ln 0.01=−4.6051...), while for 99.9% confidence one uses a "rule of 6.9."

Some readers will wonder why one cannot immediately calculate the probability of $0/n$ from the Poisson distribution: one can indeed do so, solving instead the equation $\exp(-n \cdot \text{maximum risk})=0.05$. As such readers might expect, the resulting answer of $3/n$ is the same.

If we apply this rule to our examples, and confine our inferences to the populations from which the patients derive, we can conclude with 95% confidence as follows:

EXAMPLE 1.—The maximum risk of abnormal testicular function is not greater than 21% ($3/n=3/14=21\%$) after chemotherapy. Here, the "correct" upper confidence limit based on 0/14 is 19%. Interestingly, this "rule of 14 consecutive failures" is commonly used in cancer treatment research[5] to screen out agents that are unlikely to show activity in at least one patient in five.

EXAMPLE 2.—The frequency of malformations does not exceed 2.7% in offspring of women immunized against rubella during pregnancy.

EXAMPLE 3.—The frequency of neural tube defects is not greater than 1.8% in siblings of patients with TED. Note that this 0% to 1.8% range includes the rate of 1.5 per 1,000 in the general population as well as the recurrence risk for second-degree relatives of patients with neural tube defects.

Table 2 summarizes the rule of three and tabulates the "true" risks that can be ruled out given the occurrence of zero events in series of different sizes.

Comment

Although the examples we have discussed reflect our own interests, one may observe a zero numerator in very diverse contexts: a new diagnostic test that has not yet misclassified a patient, a still-perfect surgical record, a field trial of a vaccine that uncovered no major side effects, an ophthalmology practice in which no patient with glaucoma was younger than 23 years, an airline that has never had a fatality. Thus, understanding the limits of the inferences that can be made with such an observation is important.

It is interesting that although a finding of 0 events in n observations can be interpreted in much the same statistical way as any other binomial rate, it often seems to have a qualitative impact far in excess of its quantitative meaning. A few (sometimes contradictory) clues in the literature may help to explain this: (1) People tend to ignore the size of the denominators on which rates are based.[6] For instance, a rate of 10% is given much the same credence whether it is observed in 20 or 200 cases. Presumably the same holds for the rate of 0%. (2) People tend to focus on numerators. Parents who have had genetic counseling, for example, generally view their risks in binary form: something either will or will not happen, no matter what the actual risk is.[7] A "one" in the numerator never disappears, no matter the size of the denominator. For them, it is not the odds that matter so much as the knowledge that the disorder is possible. Perhaps a zero numerator carries similar weight in that it suggests (falsely) that an event is impossible. (3) When one is faced with a theoretical risk that has not yet manifested itself, ie, when something possible has not yet happened, people may tend to expect that it cannot happen in the future. Lacking evidence that responses may be variable, they then underestimate what the real risk may be. (Interestingly, when the Marquis de Laplace, statistical consultant to the gambling royalty of Europe and Russia, was reminded circa 1800 that never once in the preceding 5,000 years had the sun failed to rise—apparently earlier data were not considered reliable—and was asked to provide odds that it would rise on the following day, he suggested odds of roughly n:1, where n was the number of days in the series, as a *fair* bet. For someone betting on sunrise who had less trust in nature, odds of $n/3$:1 would be an even safer gamble.)

We urge a reformulation of the views of a zero numerator and encourage those reporting such observations to consider the maximum risk with which their findings are compatible. To this end, the confidence interval is helpful since it translates the results of a sample not into a single number, but rather into a range that is quite likely to contain the rate characteristic of the population. Because a confidence interval may be constructed easily from a zero numerator using the "rule of three," we hope that those fortunate enough to be able to report "no problems so far" will quantify the worst or best that a group of future patients can expect.

References

1. Blatt J, Poplack DG, Sherins RJ: Testicular function in boys after chemotherapy for acute lymphoblastic leukemia. *N Engl J Med* 1981; 304:1121-1124.
2. Preblud SR, Stetler HC, Frank JA, et al: Fetal risk associated with rubella vaccine. *JAMA* 1981;246:1413-1417.
3. Baird P, MacDonald EC: Siblings of children with tracheoesophageal dysraphism. *Can Med Assoc J* 1981;125:1083-1084.
4. Rumke CL: Implications of the statement: No side effects were observed. *N Engl J Med* 1975;292:372-373.
5. Holland JF, Frei E (eds): *Cancer Medicine*, ed 2. Philadelphia, Lea & Febiger, 1982, p 535.
6. Tversky A, Kahneman D: Judgment under uncertainty: Heuristics and biases. *Science* 1974; 185:1124-1131.
7. Lippman-Hand A, Fraser FC: Genetic counseling: The postcounseling period: I. Parents' perceptions of uncertainty. *Am J Med Genet* 1979;4:51-71.