

Genetic Counseling and Recurrence Risk

Victor Matheson

First Draft: November 13, 1997

For a wide variety of genetic and developmental disorders such as chromosomal disorders and spontaneous mutations no clear cause exists for the occurrence of the disorder. Since the specific way these disorders are transmitted from the parents to the children does not follow any statistical patterns of inheritance, the genetic counselor is left with a difficult question when patients ask about the risk having another child with the same disorder. It is therefore a common practice among genetic counselors to simply quote a 1% recurrence risk to patients who have had a child with any of these types of disorders. It may at first seem that the 1% figure is of a very *ad hoc* nature and that it would be preferable to be able to give more precise risk of recurrence. However, the reality is that with the very small probabilities involved with most genetic disorders, it is not possible to obtain large enough sample sizes to be able to cite precise recurrence risks with any degree of statistical certainty. The purpose of this paper is to present a statistical model that details the reasons for not specifying more exact recurrence risks for these types of disorders and to provide support for the practice of many counselors of generally quoting a 1% recurrence risk for patients who have had one child with the disorder.

As this paper is directed towards an audience without a strong statistical background, the exact mathematics for the statistical models in this paper is left for the appendix. It is, however, worthwhile to review several terms from the language of statistics. A *Type II error* occurs when one rejects a particular scientific hypothesis (called the null hypothesis) when that hypothesis is actually true. The *significance* of a statistical test is the probability that a Type II error occurs. A *Type I error* occurs when a one does not reject a hypothesis that is false. The *power* of a statistical test is the probability that a false hypothesis is rejected which is the same as the probability of not making a Type I error.

Suppose a researcher wishes to test whether there is a higher probability of a particular disorder among patients who have previously had a child with this disorder than among the general population. To test this question, the researcher would first make the hypothesis that the prevalence of

the disorder is the same in both groups and would then collect data on subsequent births to patients who have already had one child with the disorder in an attempt to disprove that hypothesis.

However, from a statistical point of view, it is not enough to simply show that the proportion of affected children in the sample collected is higher than the proportion of affected children in regular population. Even if a researcher takes completely random samples from a population all of whom have the same risk, some samples will have more affected individuals than other samples simply by chance. A researcher will conclude that a sample comes from a population with a significantly higher risk only if the sample contains so many affected individuals that it would be very unlikely that a sample with that many affected individuals could be drawn by chance from a population with the same risk as the general population.

It is a common practice to use a 5% significance level in hypothesis testing. Thus, the researcher would conclude that the recurrence risk is higher than the risk for the general population only if the sample collected showed such a high number of affected children that the probability of collecting a random sample with so many affected children from a population that had no higher risk of the disorder than the general population was 5% or less.

As is clear from the example above, hypothesis testing is based on very conservative principles. Statisticians are reluctant to accept data that suggests that some groups may be at a higher risk for a genetic disorder unless the data clearly shows that the group must have a higher risk at a sufficiently low significance level. By choosing to reject the hypothesis that there is no increased risk of having a second child with a disorder only at very low significance levels, researchers seek to reduce the chance that a Type II error will be made. In other words, they want to minimize the chance that they will conclude that there is an increased risk of recurrence when there actually is not any increased risk for a reoccurrence of the disorder in subsequent pregnancies.

Of course, the price that researchers must pay for requiring low significance levels before concluding that there is an increased recurrence risk is that data that suggests a recurrence risk that is higher than the risk for the population as a whole may be dismissed as “statistically insignificant” because it does not meet the criteria for a specific significance level. Therefore, researchers may often make Type I errors by concluding that there is no difference between the recurrence risk and the risk

for the general population when, in fact, the recurrence risk is higher than that of the general population. This all boils down to the fact that there is a direct trade-off between the significance and the power of a statistical test and the only way to improve both is to increase the size of the research study.

If the solution to the problem of the trade-off between power and significance of a statistical test is as simple as increasing sample size, it may be tempting to conclude that this is a non-issue. Unfortunately, when dealing with events that occur relatively infrequently such as genetic disorders, the sample sizes required in order to conclude whether or not a particular disorder is associated with an increase recurrence risk at satisfactory levels of significance and power are prohibitively large.

Suppose that a certain disorder occurs in 1 out of 1,000 pregnancies. Now suppose that researcher examines 10,000 pregnancies from patients who have previously had a child with that particular disorder. Since the disorder occurs in 1 out of 1,000 pregnancies, one would expect 10 pregnancies in the sample to have the disorder. However, one would not conclude that the sample comes from a group with a statistically higher risk at the 5% significance level unless 15 or more pregnancies in the sample had the disorder. At the 1% significance level 18 pregnancies with the disorder would be required. (See Appendix A.)

Now suppose that in reality patients who have had one child with the disorder have a 50% higher chance of having a subsequent child with the disorder than the general population (i.e., the risk rises from 1 out of 1,000 pregnancies to 1 out of 667 pregnancies.) Only 53% of the time would a sample of 10,000 pregnancies from a population where 1 of 667 pregnancies are affected have the 15 or more affected pregnancies required to reject the hypothesis that the recurrence risk is no higher than the general population risk. In other words, even with a huge sample of 10,000 cases, a statistical test that checks whether the recurrence risk is higher than the risk for the general population at the 5% significance level has a power of only 53% even when the recurrence risk is truly 50% higher than the general population. In order to make a statistical study that would have a power of 95% while still maintaining a significance level of 5%, a sample size of 51,059 individuals would be needed.

Table 1 presents the sample sizes necessary in order to construct a statistical test that has a significance of 5% and a power of 95% for a disorders that range from a frequency in the general population of 1 in 1,000 to 1 in 50,000. For each frequency, the required sample size is calculated on

the assumption that the recurrence risk ranges from 1.5 times to 100 times the risk for the general population. The final column is the number of cases in total (both from first-time cases and recurrences) one would expect in the United States in a given year for each frequency based on an estimate of 4,000,000 births each year.

Table 1 Recurrence risk as multiple of general population risk

| Prevalence in General Population | 1.5x risk | 2x risk | 5x risk | 10x risk | 100x risk | Annual Cases |
|----------------------------------|-----------|----------|---------|----------|-----------|--------------|
| 1 in 1,000 live births | 51,059 | 13,936 | 1,256 | 473 | 29 | 4,000 |
| 1 in 5,000 live births | 255,323 | 69,695 | 6,294 | 2,371 | 149 | 800 |
| 1 in 10,000 live births | >500,000 | 139,392 | 12,590 | 4,742 | 299 | 400 |
| 1 in 25,000 live births | >500,000 | 348,487 | 31,477 | 11,858 | 748 | 160 |
| 1 in 50,000 live births | >500,000 | >500,000 | 62,956 | 23,718 | 1,497 | 80 |

Several results should be obvious from the chart. First, the required sample sizes rise in direct proportion with the prevalence of the disorder in the general population. Ten times as much data is required to detect a disorder that occurs in 1 in 1,000 births than 1 in 10,000 births. Next, the required sample sizes rise more than proportionally as the recurrence risk as a multiple of the general population risk falls. More than twice as much data is required to detect a disorder in which the recurrence risk is 50% higher than the risk for the general population than a disorder in which the recurrence is 100% higher than the risk for the general population.

Most importantly, for many combinations of recurrence risk and general population risk, the required sample size far exceeds the number of annual cases. The practical significance of this result is clear. The only way that a researcher can test the hypothesis that the recurrence risk is higher than the general population risk is by examining subsequent births to patients who have previously had a child with the disorder. This means that the amount of data a researcher can collect is limited by the number of subsequent births (or pregnancies) to these patients.

Suppose that a certain disorder occurs at a frequency of 1 in 10,000 births. Since only 400

cases occur nationwide in a given year, even if patients who have one child with the disorder average two subsequent pregnancies, a researcher would have to gather information on subsequent births for every single patient nationwide who had an affected child over a six year period in order to collect the 4,742 number of cases needed in order to test the hypothesis at a 5% significance and a 95% power.