Primer on Type I and Type II Errors

Statistical tests are tools that help us assess the role of chance as an explanation of patterns observed in data. The most common "pattern" of interest is how two groups compare in terms of a single outcome. After a statistical test is performed, investigators (and readers) can arrive at one of two conclusions:

1) The pattern is probably not due to chance (i.e., in common jargon, "There was a significant difference" or "The study was positive").

2) The pattern is likely due to chance (i.e., in common jargon, "There was no significant difference" or "The study was negative").

No matter how well the study is performed, either conclusion may be wrong. As shown in the Table below, a mistake about the first conclusion is labeled a type I error and a mistake about the second is labeled a type II error.

STUDY CONCLUSION	"TRUTH"	
	DIFFERENCE	NO DIFFERENCE
"Positive" study (significant difference)	True positive	Type I error
"Negative" study (no significant difference)	Type II error	True negative

Note that a type I error is only possible in a positive study, and a type II error is possible only in a negative study. Thus, this is one of the few areas of medicine where you can only make one mistake at a time.

Type I Errors

A type I error is analogous to a false-positive result during diagnostic testing: A difference is shown when in "truth" there is none. Researchers have long been concerned about making this mistake and have conventionally demanded that the probability of a type I error be less than 5%. This convention is operationalized in the familiar critical threshold for P values: P must be less than 0.05 before we conclude that a study is positive. This means we are willing to accept that in 100 positive studies, at most 5 will be due to chance alone. The probability that a type I error has occurred in a positive study is the exact P value reported. For example, if the P value is 0.001, then the probability that the study has yielded false-positive results is 1 in 1000.*

Type II Errors

A type II error is analogous to a false-negative result during diagnostic testing: No difference is shown when in "truth" there is one. Traditionally, this error has received less attention from researchers than type I error and, consequently, may occur more often. Type II errors are generally the result of a researcher studying too few participants. To avoid the error, some researchers perform a sample size calculation before beginning a study and, as part of the calculation, assert what a "true difference" is and accept that they will miss it 10% to 20% of the time (i.e., type II error rate of 0.1 or 0.2). Regardless of how a study was planned, when faced with a negative study readers must be aware of the possibility of a type II error. Determining the likelihood of such an error is not a simple calculation but a judgment.

Role of 95% CIs in Assessing Type II Errors

The best way to decide whether a type II error exists is to ask two questions: 1) Is the observed effect clinically important? and 2) To what extent does the confidence interval include clinically important effects? The more important the observed effect and the more the confidence interval includes important effects, the more likely that a type II error exists.

To gain some experience with this approach, consider the confidence intervals from three hypothetical randomized trials in the Figure. Each trial addresses the efficacy of an intervention to prevent a localized cancer from spreading. The outcome is the relative risk (RR) of metastasis (ratio of the risk in the intervention group over the risk in the control group). The interventions are not trivial, and you assert that you only consider risk reductions of greater that 10% to be clinically important. Note that each confidence interval includes 1—that is, each study is negative. There are no "significant differences" here. Which study is most likely to have a type II error?

*This statement only considers the role of chance. Readers should be aware, however, that observed patterns may also be the result of bias.
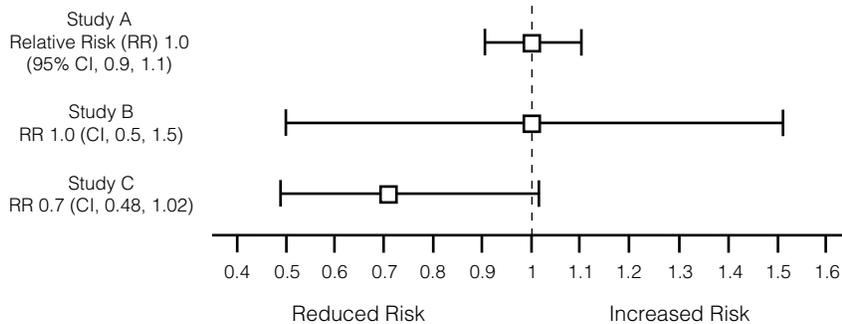
FIGURE. Role of 95% CIs in assessing type II errors.

Study A suggests that the intervention has no effect (i.e. the relative risk is 1) and is very precise (i.e., the confidence interval is narrow). You can be confident that it is not missing an important difference. In other words, you can be confident that there's no type II error.

Study B suggests that the intervention has no effect (i.e., the RR is 1) but is very imprecise (i.e., the confidence interval is wide). This study may be missing an important difference. In other words, you should be worried about type II error, but this study is just as likely to be missing an important harmful effect as an important beneficial one. A type II error is possible, and it could be in either direction.

Study C suggests that the intervention has a clinically important beneficial effect (i.e., the RR is much less than 1) and is also very imprecise. Most of the confidence interval includes clinically important beneficial effects. Consequently, a type II error is very likely. This is a study you would like to see repeated using a larger sample.