Using 2X2 Tables to Evaluate Two Alternatives

It is always difficult to choose between alternatives, even if there are only two possibilities. Statisticians use the 2X2 table to quantitatively compare choices. Researchers in many fields, from business to medicine, use such tables to assist decision making. A standard 2X2 table is shown and explained below:

Study Question	**True Status (reality)**		
Test Result (what the test indicates is reality)	Positive	Negative	Total
Positive	TP - True Positive (the test result conforms to reality)	FP - False Positive (the test result is positive but reality is negative). Also known as a Type 1 or Alpha (α) error	
Negative	FN - False Negative (the test result is negative but the reality is positive). Also known as a Type 2 or Beta (β) error	TN - True Negative (the test result conforms to reality)	
Total			

2X2 Tables in Medical Screening

Medical screening is a very common use of 2X2 tables. Suppose a research team completes a study to determine whether chest xrays (CXR) are useful screening tools for lung cancer. They could conceptualize their data analysis as follows:

Clinical Question – How well does CXR screen for lung cancer?	**Disease Status** (reality)		
Test Result (what the test indicates is reality)	Positive (the patient has lung cancer)	Negative (the patient does not have lung cancer)	Total
Positive	TP – the patient has lung cancer and the test detected it	FP – the patient does not have lung cancer but the test indicates that he does	
Negative	FN – the patient has lung cancer but the test did not detect it	TN – the patient does not have lung cancer and the test indicates that he does not have it	
Total			

After completing the study, the resultant data table may look like this:

How well does CXR screen for lung cancer?	**Disease Status** (reality)				**Result** (sample data)
Screening Test Status	Positive	Negative	Total		
Positive	TP = 315	FP = 404	719	Positive Predictive Value (PPV) = TP/(TP+FP)	PPV = 315/719 = 44%
Negative	FN = 34	TN = 247	281	Negative Predictive Value (NPV) = TN/(TN+FN)	NPV = 247/281 = 88%
Total	349	651	1000		Sens = 315/349 = 90%
Results	Sensitivity = TP/(TP+FN)	Specificity = TN/TN+FP			Spec = 247/651 = 38%

There are four key terms to understand:

1. Sensitivity is the likelihood that a test will identify a disease if it is present. It helps rule out disease when the test is negative.
2. Specificity is the likelihood that a test will not identify a disease when it is not present. It helps rule out disease when the test is positive.
3. Positive predictive value is the likelihood that if a test is positive, the person has the disease that it detects.
4. Negative predictive value is the likelihood that if a test is negative, the person does not have the disease that it detects.
5. Prevalence is the commonness of a condition in a sample or population.

We now need to apply these terms in the example above. The prevalence of lung cancer in this sample is very high, 31.5%. Such prevalence would not be found in a representative sample of the general population but might be found in population of smokers over 60 with hemoptysis.

1. Sensitivity - If patient has cancer, there is a 90% chance that the CXR will detect it.
2. Specificity – If the patient does not have cancer, there is a 62% chance that the CXR will not show lung cancer.
3. PPV - If the CXR is positive, there is a 44% chance that the patient really has lung cancer.
4. NPV - If the CXR is negative, there is an 88% chance that the patient really does not have lung cancer.
5. FNER – False negative error rate (Beta or Type 2 error) = 34/349
6. FPER – False positive error rate (Alpha or Type 1 error) = 404/651
7. LR+ - Likelihood of the outcome if the test is positive
8. LR- - Likelihood of the outcome if the test is negative
9. Prevalence – Likelihood of the disease in the population

Screening tests for hypertension, glaucoma, and many other medical conditions can be evaluated in the same way. The primary difference is that they require different tests, such as blood pressure machines, intraocular pressure sensors, blood tests and the like. Screening can save lives and money, but it can also cause harm. There are six possible outcomes of medical screening:

Six Possible Outcomes of Screening

1. Negative screen; patient does not have the disease, injury or other condition
2. Negative screen; patient has the disease, injury or other condition (false negative)
3. Positive screen; patient does not have the disease, injury or other condition (false positive)
4. Positive screen; patient has the disease, injury or other condition, but will not experience morbidity or mortality related to it in his or her lifetime
5. Positive screen; patient has the disease, injury or other condition, but treatment before the development of symptoms does not result in a longer or better life compared to treating the it when it becomes clinically apparent. The amount of time patient is considered “diseased” and is subjected to the risks and costs of treatment is lengthened.
6. Positive screen; patient has the disease, injury or other condition. Treatment before the development of symptoms lengthens life or reduces morbidity compared to treatment when the condition becomes clinically apparent

As you can see, only two out of the six possible outcomes could be considered “good”. If a screening test result is negative and the person does not have the condition of interest (option 1), the process is done. If a screening test is positive, the person has the condition of interest, and treatment at that point will enhance and prolong life (option 6), the test has served its purpose. In every other circumstance, those screened do not benefit from the screening.

2X2 Table in Evaluating Medical Therapies

2X2 tables can also evaluate medical interventions, such as drugs, vaccines, and surgeries. Suppose our intrepid research team wants to evaluate the effectiveness of influenza vaccine in a season. The conceptual table might appear as shown below:

Does Influenza Vaccine help prevent clinical influenza?	**Disease Status** (reality)		
Intervention Status (did the subject get the intervention)	Positive (the subject got clinical influenza)	Negative (the subject did not get clinical influenza)	Total
Positive (subject got the intervention)	A - Intervention Possibly Ineffective (the subject received the influenza vaccine and still developed clinical influenza)	B – Intervention Possibly Effective (the subject received the influenza vaccine and did not develop clinical influenza)	
Negative (subject did not get the intervention)	C – the subject did not receive the influenza vaccine and developed clinical influenza	D – the subject did not receive the influenza vaccine and did not develop clinical influenza	
Total			

The team completes a clinical trial with 1,000 subjects evaluating the effectiveness of influenza vaccination in a given year. The resulting 2X2 table may appear like this:

Does Influenza Vaccine help prevent clinical influenza?	**Disease Status** (reality)			**Results** (sample data)
Intervention Status	Positive	Negative	Total	The chance of a vaccinated subject getting clinical influenza (incidence of influenza in the vaccine exposed group, I_E) was 111/694 = 16%. The chance of an unvaccinated person getting clinical influenza (incidence of influenza in the vaccine unexposed group, I_N) was 235/306 =77%. The relative benefit of influenza vaccination in this study was 16/77 = 21%. Vaccinated people were about 5x less likely to get clinical influenza than unvaccinated people.
Positive	A = 111	B = 583	694	
Negative	C = 235	D = 71	306	
Total	336	664	1000	

Now we have more terms to consider:

1. Incidence (I) – The number of new cases of a disease, injury or other condition arising during a defined time period. Total Incidence (I_T) in this example is 336/1000 or 34%.
2. Prevalence (P) – How common is the disease, injury, or condition in the population?
3. Attributable risk/benefit ($AR = I_E - I_N$) - What is the incidence of a condition (good or bad) due to an exposure (harmful or helpful)? In this example, how much protection against clinical influenza is attributable to receiving the vaccine?
4. Relative risk/benefit ($RR = I_E/I_N$) - How many times more likely are exposed persons to have the outcome of interest than unexposed persons?
5. Population attributable risk (AR_P = ARxP) - What is the incidence of an outcome in the population, associated with a risk factor (such as smoking) or a protective factor (such as vaccinations)?
6. Population attributable fraction ($AF_P = AR_P/I_T$) - What fraction of disease or other condition in a population is attributable to exposure to a risk factor? What fractional decrease of disease or other condition in a population is attributable to a protective factor?

Relative Risk can also be calculated by the following: RR = $Risk_{exposed}/Risk_{unexposed}$ = [A/(A+B)]/[C/(C+D)]. If the risk of a condition is low (<5%), the RR can be approximated by the Odds Ratio (OR). The equation for the OR is OR = AD/BC. RR and OR are very useful in quantifying data and are commonly used to evaluate the relationship between risks like smoking, obesity, and drug use, and disease. A RR or OR > 1 suggests a risk, while a RR or OR < 1 suggests a benefit. In this example, the RR = 0.21 and the OR = 0.06. Because the risk is high (benefit of influenza is high), OR is not a good approximation of RR.

2X2 Tables Outside of Medicine - Screening

Looking at the effectiveness of medical screening and interventions is one of the most important uses of 2X2 tables. However, these tables are equally useful outside of medicine. Suppose that a company develops a pre-employment questionnaire to improve the quality of the people that they hire. They could evaluate the questionnaire as follows:

1. Use the questionnaire for a random selection of prospective employee interviews. There must be an intervention group (which uses the survey) and a control group (which does not).
2. The desired outcome is, of course, that the company hires people who will turn out to be good employees. However, "good employees" needs to be measureable for the study. Perhaps a reasonable surrogate outcome for "good employee" is "top 20% on performance evaluations", which may be at 3, 6 and 12 months after hiring.
3. Periodically throughout the evaluation time frame, the company would check the results.

Does hiring questionaire A (HQ_A) help company B hire people who become better employees?	**True Status** (reality)		
Intervention Status (did the prospective employee take the new survey?)	Positive (the employee ranked in the top 20% on performance evaluations)	Negative (the employee did not rank in the top 20% on performance evaluations)	Total
Positive (subject took the **HQ_A**)	A - **HQ_A** Possibly Effective	B – **HQ_A** Possibly Ineffective	
Negative (subject did not take the **HQ_A**)	C – the subject did not take **HQ_A** but ranked in the top 20% anyway	D – the subject did not take **HQ_A** and did not rank in the top 20%	
Total			

If the hiring questionnaire proved effective at helping to identify candidates who later became good employees in the pilot study, the company might expand the program and give it to all of their prospective employees. If they did, the firm would continue to monitor its effectiveness (or lack thereof).

2X2 Tables Outside of Medicine – Evaluating Interventions

Educators, whether in academia, industry, or even politics can explore the effectiveness of educational campaigns with 2X2 tables. Suppose a consulting firm was advising a medium-sized business with 60 employees on how to comply with Occupational Safety and Health (OSHA) knowledge requirements in their industry. They could follow these steps:

1. Devise a training program, including tests, to meet OSHA standards.
2. Perform a proof of concept trial.
3. Assign 10 people each to intervention and control groups. Administer a pretest to both groups.
4. Match the groups for pretest scores, experience, education, age, and other pertinent variables.
5. Train the intervention group, telling the control group to train on their own.

6. Give the posttest to both groups.
7. Compare the results using a 2X2 table.

If the training proved effective, the consulting firm would inform management and probably receive funding to train the rest of the employees. The concept 2X2 table might appear as follows:

Does educational intervention A (EI_A) help group B pass test C?	**True Status** (reality)		
Intervention Status (did the subject get the intervention)	Positive (the subject passed test C)	Negative (the subject did not pass test C)	Total
Positive (subject got **EI_A**)	A - **EI_A** Possibly Effective	B – **EI_A** Possibly Ineffective	
Negative (subject did not get **EI_A**)	C – the subject did not receive **EI_A** but passed the test anyway	D – the subject did not receive **EI_A** and did not pass the test	
Total			

The 2X2 table can be used for any binary screening, intervention or outcome. A group asking the question "Should this entity be awarded the job/grant/project" can develop a tool to measure it. One limitation is that 2X2 tables cannot be used for variables that have many possible values (such as continuous ones like blood pressure readings or income values). Though the table is useful for analysis, the data it produces can never be better than the tools that provide the data. For example, no excellence in analysis can overcome bad data from a flawed questionnaire.

Conclusion

Modern life has thousands of choices, more than ever before. Though information is everywhere, it is tough to know what the most useful information is and where to find it. Further, getting information is not enough to make the tough decisions required. Once the team gets the data that it needs, they must rigorously analyze it. For institutions without a statistician or formal data analyst, 2X2 tables are a simple way to group and study complex data. Using them can give leaders an edge in the hypercompetitive 21st century.

Please see Epidemiology & Biostats for Curious Clinicians at http://mdharrismd.com/2014/01/04/medical-lectures/ for more information.