

## EPIDEMIOLOGICAL CONCEPTS

Exposure	Outcome	
	Yes	No
Yes	a	b
No	c	d

Metric	Definition	Study Types	
		Cohort	Case-Control
Odds with exposure a/b	Odds of an event occurring if exposure occurs.	■	■
Odds w/out exposure c/d	Baseline odds of something happening (w/out exposure to condition being tested).	■	■
Risk difference (c/(c+d))-(a/(a+b))	Absolute difference in the risk between the exposed and non-exposed groups.	■	□
Relative risk (RR) (a/(a+b))/(c/(c+d))	The risk of a thing happening in an exposed group relative to the baseline risk.	■	□
Relative risk reduction 1-RR	The proportion of a baseline risk that is removed with exposure.	■	□
Odds ratio ad/cb or (a/b)/(c/d)	The odds of a thing happening in an exposed group relative to the baseline odds.	■	■
Prevalence (a+c)/(all)	Proportion of the population that have the disease.	■	□

Screening/ Test Result	Presence of Disease	Yes	No
		Positive	Negative
	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Sensitivity (TP/(TP+FN))100	Among those who are truly positive, the percent who screen/test positive.
Specificity (TN/(TN+FP))100	Among those who are truly negative, the percent who screen/test negative.
Accuracy ((TP+TN)/(all))100	Among all of the people in the 2x2 table, the percent who are identified correctly by the screen/test.
Positive Predictive Value (TP/(TP+FP))100	Among those who test positive, the percent who are true positives.
Negative Predictive Value (TN/(FN+TN))100	Among those who test negative, the percent who are true negatives.

Lead-Time Bias	The apparent lengthening of survival due to early diagnosis.
Length-Time Bias	Tendency of screening to detect larger number of cases in slowly progressing disease, but undercount diseases with rapid progression.

## STATISTICAL CONCEPTS

**Central Tendency**  
**Mean:** The sum of the scores divided by the number of cases. Use parametric statistical tests. **Median:** The score above and below which 50% of the cases fall. Use non-parametric statistical tests. **Mode:** Most frequent score.

**Variation**

**Range:** The difference between the high score and the low score. **Standard Deviation:** In normally-distributed data, 68% of cases fall within the first standard deviation above and below the mean. **Standard Error:** Standard deviation divided by the square root of the number of observations. **Confidence Interval:** Range within which the mean of 95% of repeated samples will likely fall. Mean +/- two times the standard error.

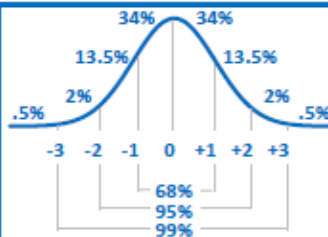
**Levels of Measurement**

Type	Definition	Statistics
Dichotomous Yes, no	Two options, qualitatively different	Percentages
Categorical Blue, green		
Ordinal Agree, disagree	Along a continuum	Percentages (mean, median and mode in certain circumstances)
Interval Temperature		
Ratio Age		
	Consistent units along continuum	Mean, median and mode
	Consistent units with a true zero	Mean, median and mode

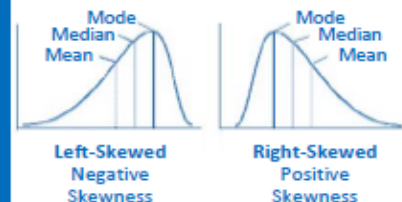
**Choosing Statistical Tests**

What You Want	Continuous	Discrete	Example
Compare different groups of study participants	Independent groups T-test (parametric); Mann Whitney U (non-parametric)	Chi-Square test (compare proportions)	Comparing the scores of boys and girls who took the same test.
Compare two points in time, the same group of subjects	Paired T-test (parametric); Wilcoxon Rank (non-parametric)	McNemar's test (dichotomous data)	Comparing the before and after scores of a group of children exposed to a certain treatment.
Compare to a benchmark/to the population	Confidence intervals	Confidence intervals	Comparing your hospital's scores to those of a group of other hospitals.
Test the association/'relatedness' of two different groups	Pearson correlation	Chi-Square test	Associating the age and height of a group of children.

**Normal Distribution**



**Skew**



Left-Skewed  
Negative  
Skewness

Right-Skewed  
Positive  
Skewness