Calculating Odds Ratio and Relative Risk in Excel: A Tutorial for Epidemiological Data Analysis

Authored by

Mohammed looti

November 8, 2025

Understanding 2x2 Contingency Tables in Epidemiology

In the rigorous fields of biostatistics and **epidemiology**, analytical studies hinge on the structured presentation of data, often relying on the use of a **2-by-2 table**, also commonly referred to as a contingency table. This foundational statistical tool is indispensable for systematically investigating the relationship between a specific exposure (or intervention) and a dichotomous outcome (such as the presence or absence of a disease or event). It provides the necessary framework to compare the frequency of an outcome between two distinct groups: typically an exposed or **treatment group** and an unexposed or **control group**. Establishing this clear organizational structure is the essential first step before researchers can proceed to accurately quantify measures of association, including the odds ratio and relative risk.

The standard 2x2 contingency table organizes observational data based on the intersection of two binary variables: **Exposure/Treatment Status** (Yes/No) and **Outcome/Event Status** (Yes/No). The four resulting cells are conventionally labeled A, B, C, and D. These labels represent specific counts of observations corresponding to the four possible combinations of exposure and outcome status. This precise categorization is critical for isolating the effect of the exposure under investigation relative to the baseline rate observed in the control population, thereby minimizing ambiguity in subsequent statistical modeling.

The illustration below visually confirms the standard arrangement of a conventional 2x2 contingency table, detailing how counts are distributed to facilitate the calculation of key epidemiological metrics:

	Event	No Event
Treatment	A	B
Control	C	D

Within this structure, Cell A represents the number of individuals who were exposed and experienced the event, while Cell B represents those exposed who did not experience the event. Conversely, Cell C represents the number of unexposed individuals who experienced the event, and Cell D represents the unexposed individuals who remained event-free. The row and column totals (A+B, C+D, A+C, B+D) are critical, as they define the total sample sizes for the exposed and unexposed populations, forming the denominators necessary for calculating risk and odds measures.

Defining the Odds Ratio (OR) and Its Calculation

The **odds ratio** (OR) stands as a paramount measure of association, especially vital when analyzing retrospective study designs, such as case-control studies, where the direct calculation of incidence rates is often impractical or unreliable. The OR is designed to quantify the strength of the relationship between an exposure and an outcome by comparing two distinct ratios: the odds of the event occurring in the exposed (treatment) group relative to the odds of the event occurring in the unexposed (control) group. Interpretation of the OR is straightforward: a value exceeding 1.0 indicates a positive association (increased odds of the outcome with exposure), while a value less than 1.0 suggests a protective factor (decreased odds).

Mathematically, the odds are fundamentally defined as the ratio of the frequency of an event happening to the frequency of the event not happening. Therefore, the odds among the exposed group are calculated as A/B, and the odds among the unexposed group are calculated as C/D. The **odds ratio** itself is the simple ratio derived from dividing the exposed odds by the unexposed odds. A key methodological advantage of the OR is its consistency; its value remains the same regardless of whether the analysis is conditioned on the row totals or the column totals, making it highly versatile in contexts like logistic regression modeling.

Utilizing the standard nomenclature (A, B, C, D) derived from the 2x2 contingency table, the formula most frequently employed for calculating the **odds ratio** relies on the cross-product of the cell frequencies. This provides a statistically sound and computationally efficient method to estimate the magnitude of association between the exposure and the outcome:

Odds ratio = (A * D) / (B * C)

Defining Relative Risk (RR) and Its Calculation

In direct contrast to the odds ratio, the **relative risk** (RR), often interchangeably termed the risk ratio, is the association measure of choice for prospective studies, such as cohort studies and randomized controlled trials. Since these designs track subjects over time, they allow for the direct calculation of incidence rates. The RR compares the incidence rate (or cumulative risk) of an outcome in the exposed group directly to the incidence rate in the unexposed group. Essentially, the relative risk answers the crucial question: what is the ratio of the **probability** of an event occurring in the treatment group compared to the **probability** of the event occurring in the control group?

The interpretation of the RR is highly intuitive and easy to communicate. For example, an RR of 2.0 indicates that the exposed population is twice as likely to experience the specific outcome compared to the unexposed population. Conversely, an RR of 0.5 suggests that the exposed group has half the risk of the outcome. This clear, direct interpretation of risk makes the **relative risk** an exceptionally powerful metric for informing public health policy, clinical decision-making, and assessing intervention efficacy.

To successfully calculate the **relative risk**, two components are required: first, the risk (proportion) of the event occurring in the exposed group (RiskExposed = A / (A + B)), and second, the risk in the unexposed group (RiskUnexposed = C / (C + D)). The ratio derived from dividing these two proportions yields the final RR value:

Relative risk = /

Practical Application: Calculating OR and RR in Excel

To solidify the theoretical definitions, we will now walk through a practical demonstration of calculating both measures using Microsoft Excel. Consider a hypothetical study on athletic training performance involving 100 basketball players. Fifty players are randomly assigned to a **new training program** (our exposed group), and the remaining 50 players are assigned to the **old training program** (our control group). The primary outcome is whether the player successfully passes a standardized skills test administered after the conclusion of the training period.

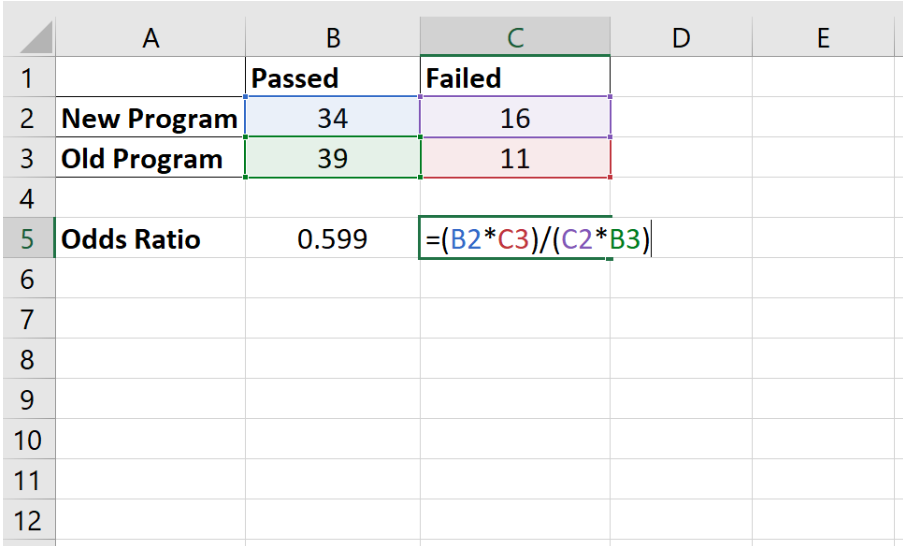
The data collected from this study is meticulously summarized in the Excel table provided below, which adheres strictly to the layout of the 2x2 contingency format, with "Passes Test" serving as the event occurrence of interest:

	A	B	C	D	E
1		**Passed**	**Failed**		
2	**New Program**	34	16		
3	**Old Program**	39	11		
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					

Based on this spreadsheet arrangement, we clearly derive our cell counts: A=34 (New Program, Passes), B=16 (New Program, Fails), C=39 (Old Program, Passes), and D=11 (Old Program, Fails). We can now efficiently compute the **odds ratio** by applying the cross-product formula: (A * D) / (B * C). Substituting the specific numerical values from the table yields the expression: (34 *

11) / (16 * 39).

Executing the calculation, the resulting **odds ratio** is determined to be: 374 / 624 ≈ **0.599**. The utility of Excel lies in its capacity for immediate verification and precise calculation of this primary measure of association, which already suggests a potential protective or detrimental effect depending on the context.



	A	B	C	D	E
1		**Passed**	**Failed**		
2	**New Program**	34	16		
3	**Old Program**	39	11		
4					
5	**Odds Ratio**	0.599	=(B2*C3)/(C2*B3)		
6					
7					
8					
9					
10					
11					
12					

Following the OR, we proceed to calculate the **relative risk**. This procedure first necessitates calculating the success probability for each training group: The risk for the New Program (RiskExposed) is 34 successes divided by 50 total players, resulting in 0.68. The risk for the Old Program (RiskUnexposed) is 39 successes divided by 50 total players, resulting in 0.78. The relative risk is the final ratio of these probabilities: 0.68 / 0.78.

The final **relative risk** calculation yields a value of approximately **0.872**. Both the OR (0.599) and the RR (0.872) consistently indicate a similar trend: the new training regimen is associated with a lower success rate compared to the established, old program, prompting further inquiry into its effectiveness.

	A	B	C	D	E
1		**Passed**	**Failed**		
2	**New Program**	34	16		
3	**Old Program**	39	11		
4					
5	**Odds Ratio**	0.599			
6	**Relative Risk**	0.872	=(B2/(B2+C2)) / (B3/(B3+C3))		
7					
8					
9					
10					
11					
12					

Interpreting the Results and Magnitude of Effect

The derived point estimates--OR = 0.599 and RR = 0.872--are critical data points requiring careful interpretation to draw statistically sound and actionable conclusions. Since both calculated measures fall below 1.0, they unanimously suggest that participation in the new training program results in a reduced likelihood of successfully passing the skills test when compared to the established old program.

For the **odds ratio** (OR = 0.599), the precise interpretation states that the odds of a player achieving a passing score while utilizing the new program are only **0.599 times the odds** of passing the test using the old program. To quantify the magnitude of this reduction, we calculate the difference from unity: (1 - 0.599) = 0.401. This implies a significant 40.1% reduction in the odds of success associated with the implementation of the new training regimen. Such a considerable finding would immediately necessitate a deep investigation into the design and efficacy of the ostensibly less effective new program.

For the **relative risk** (RR = 0.872), the interpretation is anchored directly in risk and probability. This result means that a player enrolled in the new program has 87.2% of the probability (or risk) of passing the test compared to a player enrolled in the old program. This can be verified by reviewing the underlying calculated probabilities:

Probability of passing under the new program = 34 / 50 = **68%**
Probability of passing under the old program = 39 / 50 = **78%**

The relative risk calculation (0.872) therefore accurately reflects the ratio between these two observed absolute risks (0.68 / 0.78), powerfully confirming the potentially detrimental association of the new program in terms of observed outcome frequency.

Calculating Confidence Intervals for the Odds Ratio

While point estimates (OR and RR) provide a measure of effect magnitude, they do not account for inherent sampling variability. To assess the precision and statistical significance of our findings, it is imperative to compute **confidence intervals** (C.I.). A 95% C.I. defines the plausible range of values within which the true population parameter is expected to lie 95% of the time. Crucially, if this interval encompasses the null value (1.0 for both OR and RR), the observed result is conventionally deemed not statistically significant at the 95% level.

Calculating the 95% **confidence interval** for the **odds ratio** requires a crucial transformation step. The OR must first be converted using the natural logarithm (ln(OR)) because the distribution of the log odds ratio approximates a normal distribution, which is necessary for standard error calculations. We must then compute the Standard Error of the log odds ratio, abbreviated as SE(ln(OR)).

The standard error is calculated directly from the cell counts (A, B, C, D). Once the interval limits are calculated on the log scale, they are exponentiated (e raised to the power of the limit) back to the original odds ratio scale. The general formula for the 95% C.I. for the odds ratio is defined as:

95% C.I. for odds ratio =

Where the standard error of the log odds ratio, SE(ln(OR)), is calculated using the reciprocal sum of the cell frequencies:

SE(ln(OR)) = √1/A + 1/B + 1/C + 1/D

Applying this comprehensive calculation to our basketball example (where OR = 0.599), the resulting 95% C.I. for the odds ratio is calculated to be **(0.245, 1.467)**. Since this interval clearly contains the null value of 1.0, we are compelled to conclude that we do not possess sufficient statistical evidence to definitively state that the new program's odds of success are significantly different from the odds associated with the old program. The complete Excel implementation of this complex formula set is illustrated below:

	A	B	C	D	E
1		**Passed**	**Failed**		
2	**New Program**	34	16		
3	**Old Program**	39	11		
4					
5	**Odds Ratio**	0.599			
6	**Relative Risk**	0.872			
7					
8	**SE(ln(OR))**	0.457	=SQRT(1/B2+1/C2+1/B3+1/C3)		
9	**95% C.I. for Odds Ratio**	0.245	=EXP(LN(B5)-1.96*B8)		
10		1.467	=EXP(LN(B5)+1.96*B8)		
11					
12					
13					
14					
15					
16					

Calculating Confidence Intervals for Relative Risk

Obtaining a 95% **confidence interval** (C.I.) for the **relative risk** is equally critical for validating the robustness of the study's conclusions. Similar to the OR procedure, the RR must undergo a logarithmic transformation (ln(RR)) before interval calculation to ensure the estimate is calculated on a scale that adheres to the assumptions of normality, after which it is transformed back to the original risk ratio scale.

The standard error calculation for the log relative risk, SE(ln(RR)), requires a slightly specialized formula tailored specifically to the risk calculation methodology. This formula carefully accounts for the total population sizes within both the exposed (A+B) and unexposed (C+D) groups.

The formula for establishing the 95% C.I. for the **relative risk** is defined as follows, relying on the exponentiation of the log-transformed limits:

95% C.I. for relative risk = exp(ln(RR) - 1.96*SE(ln(RR))) to exp(ln(RR) + 1.96*SE(ln(RR)))

Where the standard error of the log relative risk is precisely calculated using the following formula:

SE(ln(RR)) = √1/A + 1/C - 1/(A+B) - 1/(C+D)

When we meticulously apply this calculation to our athletic performance data (where RR = 0.872), the resulting 95% C.I. for the relative risk spans the range of **(0.685, 1.109)**. Because this calculated range encompasses the critical null value of 1.0, we draw the identical statistical conclusion reached with the odds ratio: the observed difference in passing rates is not statistically significant. This outcome suggests that the difference observed between the new and old programs may plausibly be attributed merely to random chance sampling variability. The full Excel computation demonstrating these steps is visualized below:

	A	B	C	D	E	F
1		**Passed**	**Failed**			
2	**New Program**	34	16			
3	**Old Program**	39	11			
4						
5	**Odds Ratio**	0.599				
6	**Relative Risk**	0.872				
7						
8	**SE(ln(OR))**	0.457				
9	**95% C.I. for Odds Ratio**	0.245				
10		1.467				
11						
12	**SE(ln(RR))**	0.123	=SQRT(1/B2+1/B3-1/(B2+C2)-1/(B3+C3))			
13	**95% C.I. for Relative Risk**	0.685	=EXP(LN(B6)-1.96*B12)			
14		1.109	=EXP(LN(B6)+1.96*B12)			
15						
16						
17						
18						

Additional Resources for Epidemiological Analysis

To further advance your proficiency in interpreting and applying these crucial epidemiological measures, particularly when conducting data analysis in software like Excel, the following supplementary resources and related tutorials are highly recommended for providing deeper context and methodological insight:

Detailed conceptual guides that meticulously explain the fundamental differences and appropriate uses of the **Odds Ratio** versus the **Relative Risk** in various study designs.
Advanced tutorials focusing on the rigorous interpretation of measures of effect when the null hypothesis value (1.0) is contained within the computed 95% confidence interval.
Resources covering specialized statistical methods for analyzing stratified 2x2 data, such as the widely utilized Mantel-Haenszel technique, which controls for potential confounding variables.