

A One-Sided Fisher's Exact Test: A Tail of Clinical Worsening

Bill Coar, Axio Research

ABSTRACT

In many disease areas, therapeutic intervention is designed to prevent clinical worsening of the disease, thereby increasing the quality and length of life for patients with life threatening diseases. While this research hypothesis may be simple to state in non-statistical terms, there are many endpoints that can be used to provide insight to this question.

One of the common approaches to address this question is to simply ask if a patient experiences clinical worsening after some fixed exposure to a new therapy. For diseases that rapidly progress, intervention of a new therapy should delay clinical progression for the drug to be deemed efficacious. This suggests that the proportion of patients who experience clinical worsening while on active study drug should be lower than the proportion of patients experiencing clinical worsening in a control group. There are multiple ways to demonstrate this with a statistical test.

The use of a Fisher's Exact Test will be assumed for this presentation. The focus, however, is not on the test itself, but rather on the structure of the input dataset used with PROC FREQ to obtain the inferential statistics (i.e., p-value). This is critical when the research hypothesis is one-sided in nature. In a one-sided setting, the structure and order of the treatment and response variables is essential to understanding if a right-tailed or a left-tailed p-value is appropriate so that it is consistent with the underlying research hypotheses.

This presentation will use a placebo-controlled clinical study to show how the order of the treatment and response variables impacts the statistical hypotheses associated with a one-sided Fisher's Exact Test. While the hypotheses themselves may differ, the research question remains the same...Does the study drug reduce the incidence of clinical worsening?

INTRODUCTION

Absolute differences, relative risk, and odds ratios are different measures that can be used to compare event rates (proportions). When the sample sizes are small, large-sample asymptotic distributions may not be applicable when making inference. Exact distributions play a critical role to make inferences in these situations. The remainder of this discussion focuses on performing a Fisher's Exact Test using data from a 2x2 cross tabulation, and how the sort order of the underlying data impacts the statistical hypothesis to be tested. Discussion of the statistical theory is left to the reader (see [1]).

MOTIVATING EXAMPLE

A clinical trial is being conducted to see if a new treatment can improve the general well-being of a patients with a degenerative disease. Data on many endpoints are captured, and many questions will be answered. One question of particular interest is related to clinical worsening. Clinical worsening is a dichotomous endpoint (yes or no) that is usually a function of several parameters associated with the underlying disease. The definition is usually specific to individual disease areas, and is outside the scope of this discussion. For the purpose of this presentation, the reader only needs to view clinical worsening as a measure of progression of disease or worsening of symptoms. If the disease appears to have progressed or symptoms are worsened, then a patient is said to have clinical worsening. If a new drug is effective, then one would expect fewer patients in the Treatment arm to experience clinical worsening.

Data were summarized and trial results reviewed by study team members. Example results are presented below in Table 1.

Table 1: Study Results for Clinical Worsening

	Treatment N=30	Placebo N=10
Any Clinical Worsening	5 (16.7%)	0 (0%)
P-value (Fishers Exact Test)	1.00	

During the review process, one reviewer questioned the results. With 5 vs 0, shouldn't the p-value be < 1.0? In order to make sense of both the question and the results, we introduce the odds ratio as a precursor to a Fisher's Exact Test.

ODDS RATIO

The odds ratio is a ratio of odds between two groups and is expressed as a function of cell probabilities. The 2x2 table of probabilities is written as:

Table 2: 2x2 Cross Tabulation of Probabilities

	Success	Failure
Treatment	π_1	$1-\pi_1$
Placebo	π_2	$1-\pi_2$

Note that the first column defines a success (clinical worsening in our initial example), which drives the following discussion. The odds of success are defined as:

$$odds_i = \frac{\pi_i}{1 - \pi_i}$$

where $i=1$ for treatment and $i=2$ for placebo. If the probability of success > probability of failure, then the odds are greater than 1. Conversely, if the probability of success < probability of failure then the odds are less than 1. Odds are a way of describing how much more (or less) likely we are to observe a success rather than a failure. For example, odds=4 suggest a success is 4x more likely than failure. Odds=0.25 suggest a success is 4x less likely than failure.

An odds ratio is a ratio of odds, and is defined by:

$$\theta = \frac{\pi_1 / (1 - \pi_1)}{\pi_2 / (1 - \pi_2)}$$

If the odds of success are higher in the treatment arm, the odds ratio $\theta > 1$. If the odds of success are the same, then $\theta = 1$, and if the odds of success are higher in the placebo arm then $\theta < 1$. Note the description is in terms of odds of success since that is defined in the first column. Various hypotheses that can be tested are described below.

$H_0: \theta = 1$
 $H_a: \theta \neq 1$
 Two sided test

Odds of success differ

$H_0: \theta = 1$
 $H_a: \theta > 1$
 Right sided test

Odds of success are greater in treatment

$H_0: \theta = 1$
 $H_a: \theta < 1$
 Left sided test

Odds of success are less in treatment

These hypotheses are dependent on the manner in which cell (1,1) is defined. In our discussion, we defined **success in treatment arm** to be cell (1,1). The definition of cell (1,1) is extremely critical when a one-tailed test is desired.

FISHER'S EXACT TEST

When sample sizes are small, it is more desirable to use exact tests rather than approximations based on large sample distributions. Fisher's Exact is an exact test associated with a 2x2 table. With a 2x2 table, a sampling distribution for cell counts is defined by the set of 2x2 tables having the same row and column totals as the observed data. For fixed row and column totals, the value in cell (1,1) determines the other 3 cell counts. P-values are based on hypergeometric probabilities associated with the count in cell (1,1). For a more comprehensive discussion, see [1].

In all of the following examples, the research hypothesis states: Does treatment reduce clinical worsening? The statistical hypotheses in each example may need to be re-stated to accommodate the data structure, but the research hypothesis remains unchanged. Since interest is in reducing clinical worsening, the test is single sided (ie, one-tailed).

All SAS® code associated with the discussion can be found in Appendix 1.

The 2x2 table with the added row/column totals associated with our motivating example is given by Example 1:

	Clinical Worsening	No Clinical Worsening	Total
Treatment	5	25	30
Placebo	0	10	10
Total	5	35	40

Example 1

To achieve this table in SAS, one needs to rely on the sorting of the treatment group variable and outcome variable. We use numeric a treatment group (1=treatment, 2=placebo) and numeric outcome variable (1=clinical worsening, 2=no clinical worsening) to achieve this specific sort.

In this example with clinical worsening in the treatment arm in cell (1,1), the three possible hypotheses (with the desired hypothesis in red) to be tested are:

- Two tailed: odds of clinical worsening differ between treatment and placebo
- **Left tailed: odds of clinical worsening are less for patients on treatment**
- Right tailed: odds of clinical worsening are greater for patients on treatment

Recall that our interest is in a one-tailed test. We select a left tailed test or a right tailed test that is consistent with the research hypothesis, though we rely on a statistical analysis plan and/or statistician for guidance as needed. With this 2x2 table, the left tailed test is clearly consistent with the research hypothesis. There is no logic to considering a right tailed test since it contradicts the research hypothesis. The output from PROC FREQ using data from Example 1 is shown in Output 1.

Fisher's Exact Test	
Cell (1,1) Frequency (F)	5
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	0.2166
Table Probability (P)	0.2166
Two-sided Pr <= P	0.3059

Output 1

Suppose the sort order of the treatment group was different such that the Placebo group was first (as in the case where treatment=1 and placebo=0).

	Clinical Worsening	No Clinical Worsening	Total
Placebo	0	10	10
Treatment	5	25	30
Total	5	35	40

Example 2

In this example with clinical worsening in the placebo arm in cell (1,1), the two hypotheses that are one tailed are:

- Left tailed: odds of clinical worsening are less for patients on placebo
- Right tailed: odds of clinical worsening are greater for patients on placebo

The output from PROC FREQ using data from Example 2 is shown in Output 2.

Fisher's Exact Test	
Cell (1,1) Frequency (F)	0
Left-sided Pr <= F	0.2166
Right-sided Pr >= F	1.0000
Table Probability (P)	0.2166
Two-sided Pr <= P	0.3059

Output 2

It is clear that changing the sort order of the treatment group variable changes the tail of the test. While we have not changed the research hypothesis, knowledge of the sort order (and cell (1,1)) is needed to determine if the right-sided or left-sided p-value is to be reported. Note that the statistical hypotheses for the right and left tailed tests are defined by the treatment group in the first row and the outcome in the first

column.

Suppose the sort order of the outcome variable differed as well, such that 1=clinical worsening and 0=no clinical worsening. This results in the table for Example 3.

	No Clinical Worsening	Clinical Worsening	Total
Placebo	10	0	10
Treatment	25	5	30
Total	35	5	40

Example 3

In this example with no clinical worsening in the placebo arm in cell (1,1), the two hypotheses consistent with the research hypotheses are:

- **Left tailed: odds of no clinical worsening are less for patients on placebo**
- Right tailed: odds of no clinical worsening are greater for patients on placebo

If the drug was effective, then the probability of no clinical worsening should be greater for treatment than placebo. This implies the odds of no clinical worsening are less for placebo, hence it is consistent with the original research hypothesis.

The output from PROC FREQ using data from Example 3 is shown in Output 3.

Fisher's Exact Test	
Cell (1,1) Frequency (F)	10
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	0.2166
Table Probability (P)	0.2166
Two-sided Pr <= P	0.3059

Output 3

In the last example, suppose the order of the treatment group variable is changed resulting in the following table in Example 4:

	No Clinical Worsening	Clinical Worsening	Total
Treatment	25	5	30
Placebo	10	0	10
Total	35	5	40

Example 4

Example 4 defines cell (1,1) as the number of patients on treatment without clinical worsening. Cell (1,1) defines a success as someone on treatment without clinical worsening, which may be more intuitive and is more consistent with the research hypothesis. In this setting, the research hypothesis suggests the the probability of no clinical worsening should be greater in the treatment arm, so the one-tailed hypotheses that could be considered are:

- Left tailed: odds of no clinical worsening are less for patients on treatment
- **Right tailed: odds of no clinical worsening are greater for patients on treatment**

As stated earlier, if the drug was effective then the probability of no clinical worsening should be greater for treatment than placebo.

The output from PROC FREQ using data from Example 4 is shown in Output 4.

Fisher's Exact Test	
Cell (1,1) Frequency (F)	25
Left-sided Pr <= F	0.2166
Right-sided Pr >= F	1.0000
Table Probability (P)	0.2166
Two-sided Pr <= P	0.3059

Output 4

In all four examples, the research hypothesis remained the same. The statistical hypothesis, which is defined by cell (1,1), was adapted to accommodate the various sort orders. Once the statistical hypothesis is determined, it is easy to decide if a left-sided p-value or right-sided p-value is desirable.

CONCLUDING REMARKS

Does the new treatment reduce clinical worsening? This statement itself suggests a one-tailed test. Ironically, the study observed an increased rate of clinical worsening in the experimental treatment arm. Since this is contradictory to the research hypothesis, a p-value=1.0 is not surprising.

The examples above clearly demonstrate that sort order matters when a one-tailed test is desired. They also show that the statistical hypotheses may be worded in terms of either a success or failure. While the sort determines the contents of cell (1,1), a statistical hypothesis consistent with the research hypothesis is obtainable. Knowing the statistical hypothesis associated with cell (1,1) then allows you to easily determine if a right tailed test or a left tailed test is desired.

REFERENCES

[1] Agresti, A. (1996) "An Introduction to Categorical Data Analysis", Wiley

ACKNOWLEDGMENTS

The author would like to thank Emily Woolley, Marcia Craib and Darrin Despain for their careful review of this paper. The author would also like to acknowledge the many programmers publishing information on the internet, and apologizes to those not properly referenced above, citing the inherent nature of internet searching which makes it difficult to track so many individual ideas.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

William Coar, PhD
Biostatistician/Director of Statistical Consulting
Axio Research, LLC
Seattle, WA 98121
Email: williamc@axioresearch.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

APPENDIX 1

```
ods rtf file="fisher.rtf" style=htmlblue;
```

```
title1 "C(1,1) to be (trt=1(C),outcome=1(CW)), C(2,1)=(trt=2(P),outcome=1(CW))";  
title2 " C(1,2)=(trt=1(C),outcome=2(NO CW), C(2,2)=(trt=2(P),outcome=2(NO CW))";  
title3 "Looking for Odds of CW in TRT=1 to be less than Odds of CW in TRT=2";  
title4 "Need a left tail test";
```

```
data z;  
    trt=1;outcome=1;count=5;output;  
    trt=1;outcome=2;count=25; output;  
    trt=2;outcome=1;count=0;output;  
    trt=2;outcome=2;count=10; output;  
run;
```

```
proc freq data=z;  
tables trt*outcome / chisq exact;  
weight count;  
run;
```

```
title1 "C(1,1) to be (trt=0(P),outcome=1(CW)), C(2,1)=(trt=1(C),outcome=1(CW))";  
title2 " C(1,2)=(trt=0(P),outcome=2(NO CW), C(2,2)=(trt=1(C),outcome=2(NO CW))";  
title3 "Looking for Odds of CW in TRT=0 to be greater than Odds of CW in TRT=1";  
title4 "Need a right tail test";
```

```
data z;  
    trt=1;outcome=1;count=5;output;  
    trt=1;outcome=2;count=25; output;  
    trt=0;outcome=1;count=0;output;  
    trt=0;outcome=2;count=10; output;  
run;
```

```

proc freq data=z;
tables trt*outcome / chisq exact;
weight count;
run;

title1 "C(1,1) to be (trt=0(P),outcome=0(NO CW)), C(2,1)=(trt=1(C),outcome=0(NO CW))";
title2 " C(1,2)=(trt=0(P),outcome=1(CW), C(2,2)=(trt=1(C),outcome=1(CW))";
title3 "Looking for Odds of NO CW in TRT=0 to be less than Odds of NO CW in TRT=1";
title4 "Need a left tail test";

data z;
  trt=1;outcome=1;count=5;output;
  trt=1;outcome=0;count=25; output;
  trt=0;outcome=1;count=0;output;
  trt=0;outcome=0;count=10; output;
run;

proc freq data=z;
tables trt*outcome / chisq exact;
weight count;
run;

title1 "C(1,1) to be (trt=1(C),outcome=0(NO CW)), C(2,1)=(trt=2(P),outcome=0(NO CW))";
title2 " C(1,2)=(trt=1(C),outcome=1(CW), C(2,2)=(trt=2(P),outcome=1(CW))";
title3 "Looking for Odds of NO CW in TRT=1 to be greater than Odds of NO CW in TRT=2";
title4 "Need a right tail test";

data z;
  trt=1;outcome=1;count=5;output;
  trt=1;outcome=0;count=25; output;
  trt=2;outcome=1;count=0;output;
  trt=2;outcome=0;count=10; output;
run;

```