CFDA MRI Guidance sample size formulae issue

## Abstract

CFDA MRI guidance provides a sample size calculation formulae for MRI clinical trial which aims to compare image quality with a target value. However, CFDA guidance has a mistake in cutting α from 0.05 to 0.025 when use one-tailed test. This article aims to demonstrate this CFDA guidance mistake in details and discuss the impact of this mistake on the sample size and statistical power.

## 1. China CFDA Guidance

The CFDA guidance - 《Medical Magnetic Resonance Imaging System Registration Technical Review Guidance》 provides samle size calculation formulae for single group target value method in part 4.1. Below part 1.1 is extracted from CFDA part 4.1 directly without any modification.

#### 1.1 CFDA guidance part 4.1: Sample size required for single group target value method

According to the clinical requirements, the image quality index for clinical diagnosis must not be less than 75% (target value) (considering that MR images are largely affected by patient’s cooperation, therefore the target value is set at 75%). Assuming that the image quality index is 90%, then if the significance level (two tails) is 0.05, test effectiveness is 80%, considering 10% failure rate, according to principles of statistics, the calculated sample size for each anatomy is a minimum of 60 test human subjects. Corresponding sample size formula:

Where: is the Expected efficacy value of the corresponding test group is the corresponding target efficacy value μ is the quantile of the standard normal distribution. α is the Type I error, set to 0.025. β is the Type II error, set to 0.2 Each manufacturer may calculate the sample size based on submitted product image quality index, but it must comply with the above minimum sample size requirements

#### 1.2 CFDA guidance mistake?

In above part 4.1 in CFRA MRI guidance, we can find mistake in the significance level value (i.e., type I error α ). It is clear that the hypothesis test in the guidance is directional to test whether image quality rate is not smaller than target value 0.75. Thus, one-tailed hypothesis test is adopted:

With a one-tailed test, the alternative hypothesis is directional, whereas with a two-tailed test it is nondirectional. Here, we want to test if the image quality index for clinical diagnosis not less than 75% (target value), so the alternative hypothesis is directional and the one-tailed test is adopted. However, the guidance then mentions that “then if the significance level (two tails) is 0.05”, which is a mistake because we should use one-tail significance 0.05 here. Then, in the sample size calculation formulae, the guidance set the α to 0.025, which is also a mistake because the guidance cuts the 0.05 half to 0.025. The correct method is listed as below: α shoul be set 0.05 both in the one-tailed test and two-tailed test. The difference is in the calculation of sample size: is used in one-tailed test while is used in two-tailed test. Once one-tailed test adopted, there is no need to cut α half.

In one-tailed test, we should set the 0.05 to the one-tailed significance level instead of cutting 0.05 half. If we cut 0.05 half, we actually drop into the two-tailed test thinking when use the one-tailed test. It actually becomes very strict for one-tailed test if 0.05 is cut half. I guess CFDA provides the sample size calculation for this one-tailed hypothesis test in this guidance based on the most strict principle (using the two-tailed test thinking). But I would think this principle is inappropriate since the one-tailed test is determined with no controvercy for this MRI assessment. When there is contorvercy on choice of one-tailed or two-tailed test or we are uncertain if the directional effect size is the only possiblity (i.e., the opposite effect size is impossible), the statistical principle is to choose two-tailed test conservatively. However, once one-tailed test adopted, there is no valid justification to cut α half. Although CFDA sample size calculation formulae may be a mistake in cutting α to 0.025, we should also follow this formulae to calculate sample size in practise before the guidance may be updated in future.

## 2. Sample size calculation

Using the wrong significance level 0.025 for one-tailed test, what impact does it have on the sample size? The answer is sample size is larger with significance level 0.025 than the sample size with significance level 0.05. Why? Let us see below formulaes and R code to illustrate the calculated sample size using the example CFDA provided in guidance.

### 2.1 One-Tailed test sample size:

Below is the function used to determine the sample size

samplesize\_one <- function(a, b, p0, pt){  
   
n <- ( qnorm(1-a)\*sqrt(p0\*(1-p0))   
 + qnorm(1-b)\*sqrt(pt\*(1-pt)) )\*\*2 / (pt-p0)\*\*2  
print (n)  
}  
  
samplesize\_one(0.05, 0.2, 0.75, 0.90)

## [1] 41.36453

### 2.2 Two-Tailed test sample size:

samplesize\_two <- function(a, b, p0, pt){  
   
n <- ( qnorm(1-a/2)\*sqrt(p0\*(1-p0))  
 + qnorm(1-b)\*sqrt(pt\*(1-pt)) )\*\*2 / (pt-p0)\*\*2  
print (n)  
}  
  
samplesize\_two(0.05, 0.2, 0.75, 0.90)

## [1] 53.89279

Open the website <https://www.stat.ubc.ca/~rollin/stats/ssize/b1.html>, you can get sample size number by simply entering values for , , α, β, and choosing one-tailed or two-tailed test. Above two R code functions can get the equal sample size number with this website tool.

### 2.3 Summary

From above two formulaes, we can find clearly that is used in one-tailed test sample size while is used in two-tailed test sample size. Set α to 0.05, is 1.64 and is 1.96. This shows that two-tailed test need larger sample size than one-tailed test with the same conditions. Put in another way, under the same conditions one-tailed test has the much larger statistical power than two-tailed test.

a<-0.05   
qnorm(1-a)

## [1] 1.644854

qnorm(1-a/2)

## [1] 1.959964

## 3. Discussion

In fact, the CFDA guidance provides one-tailed test for the MRI coil image quality assessment (not less than target value 0.75). But the guidance cuts the α from 0.05 to 0.025 in calculating the sample size without any valid justification. In choice of one-tailed or two-tailed test, we should keep conservative especially when we are uncertain if the opposite direction may happen in the experiment. However, if we have strong belief that the only one direction can happen in the experiment, just like what we stronly believe that our MR coil image quality is not less than target value 0.75, we can adopt the one-tailed test. Once one-tailed test adopted, there is no need to cut α half. This will descrease the statistical power or increase the sample size. So I guess CFDA guidance is a mistake in cutting α half which is more strict for our medical device sponsor.

## 4. Discussion

Once one-tailed test is adopted in the clinical trial given that we have high belief that the only one effect size direction can happen in the clinical trial, there is no need and no any valid justification to cut type I error α half just because it is one-tailed test.

## References:

1. <https://www.itl.nist.gov/div898/handbook/prc/section2/prc242.htm>
2. <https://www.stat.ubc.ca/~rollin/stats/ssize/b1.html>
3. Bernard Rosner’s 《Fundamentals of Biostatistics》 Chapter 7 - Hypothesis Testing: One-Sample Inference
4. Shein-Chung Chow’s 《Sample Size Calculations in Clinical Research》 Chapter 2.2 One-Sided Test Versus Two-Sided Test
5. Braver,S.L. On splitting the tails unequally: a new perspective on one-versus two tailed tests. Educational and Psychological meansurement, 1975, 35, 283-301.
6. Schaffer,J.P. Directional statistical hypotheses and comparisons among means. Psychological Bulletin, 1972, 77, 195-197.
7. Good, I.J. One tail versus two-tails, and the harmonic-mean rule of thumb. Journal of statistical computation and stimulation, 1984, 19 (2), 174-176.
8. David, B.P. One- Versus Two-Tailed Hypothesis Tests in Contemoporary educational research. Educational Researcher, 1991, 20 (9), 13-17.