Statistical Assurance in SAS: An Introduction and User's Guide

Jonathan L. Moscovici, IQVIA, Montreal, QC;
Milena Kurtinecz, GSK, Philadelphia, USA;

ABSTRACT

A long-standing concept in the planning of clinical trials has been that of statistical power. However, traditional calculation of power implies requires specification of several unknown quantities such as measured treatment effect between groups. In this way, the power obtained is a conditional quantity that can be misleading if the underlying assumptions in the calculation are inaccurate. This leads to a miscalculation of the required sample size and other important trial determinants. Statistical assurance is a method of calculating the unconditional probability of success of a trial by assigning probability distributions to unknown parameters such as treatment effect, rather than just one "best guess" estimate. These assigned probability distributions can be based on expert opinion, available pilot data and clinical considerations. Although not a new concept, statistical assurance adoption has been slow in the biopharmaceutical community partly due to the lack of software implementation. This paper will introduce the basic concepts, clinical cases and implementations in SAS. We focus on statistical significance as a metric of success, but this can also be defined in terms of achieving a target clinical profile.

INTRODUCTION

The notion of statistical power is widely used in clinical trials to calculate the required number of patients or events (sample size) in order to have a reasonable chance of detecting a treatment effect, if one exists (Chen et al., 2018). For example, if you enroll a certain number of patients such that your trial has a power of 80%, this can be interpreted as there being enough patients to expect an 80% chance of finding a statistically significant improvement in outcome between your placebo and treatment group, given that the drug is effective. The idea here is that with a sufficiently high power (i.e. number of subjects), the clinical trial would not have missed an opportunity to prove the benefits of drug that actually works. For example, a highly effective drug can potentially be placed in a trial with too few patients to properly prove its efficacy. This would be a waste of time, money, and patient trust. The concepts of statistical power and sample size calculation exist to avoid this scenario. The traditional "frequentist" statistical power is computed conditionally based on assumptions such as:

1- The minimal treatment effect of interest (Do we want to detect a difference of at least 5%? 10%? 20%? 50%?)

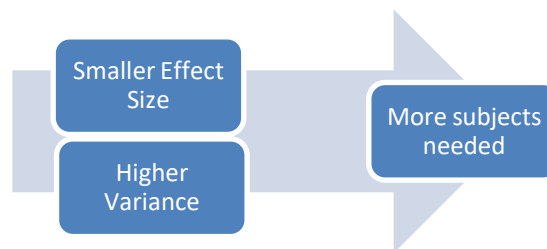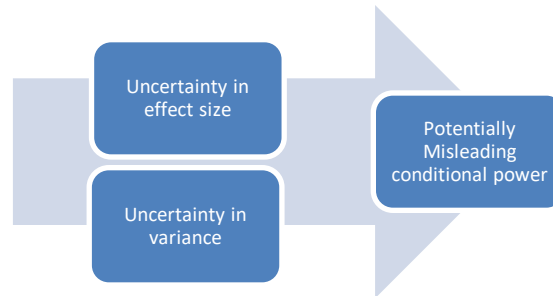2- Variability (What is the population variance or standard deviation of the outcome we are measuring?)

**Figure 1: Relationship between Assumptions and Sample Size**

Figure 2: Uncertainty in sample size assumptions

For the first assumption, one would expect that the smaller the difference to detect, the subtler the effects are, and so the more patients would be needed to provide evidence of efficacy (i.e. more accuracy is needed). For the second assumption, the higher the variability of the data, the more patients would be needed in order to provide sufficient evidence of efficacy.

During the traditional sample size calculation process, the assumptions on items 1 and 2 can have a drastic effect on the calculated sample size (Chen et al., 2016). However, it is not always possible to make realistic assumptions on these quantities, which can ultimately jeopardize the probability of a trial's success.
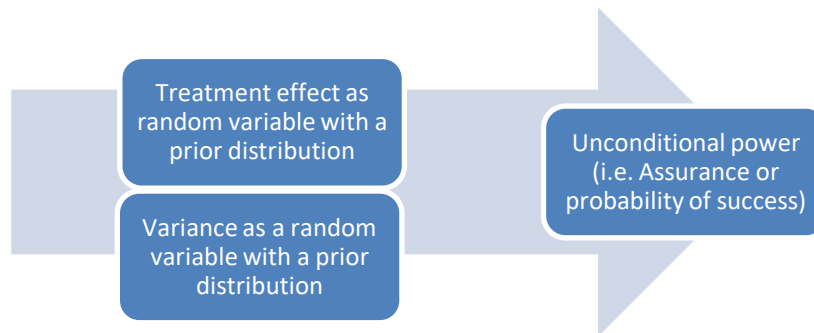
Figure 3. Statistical assurance
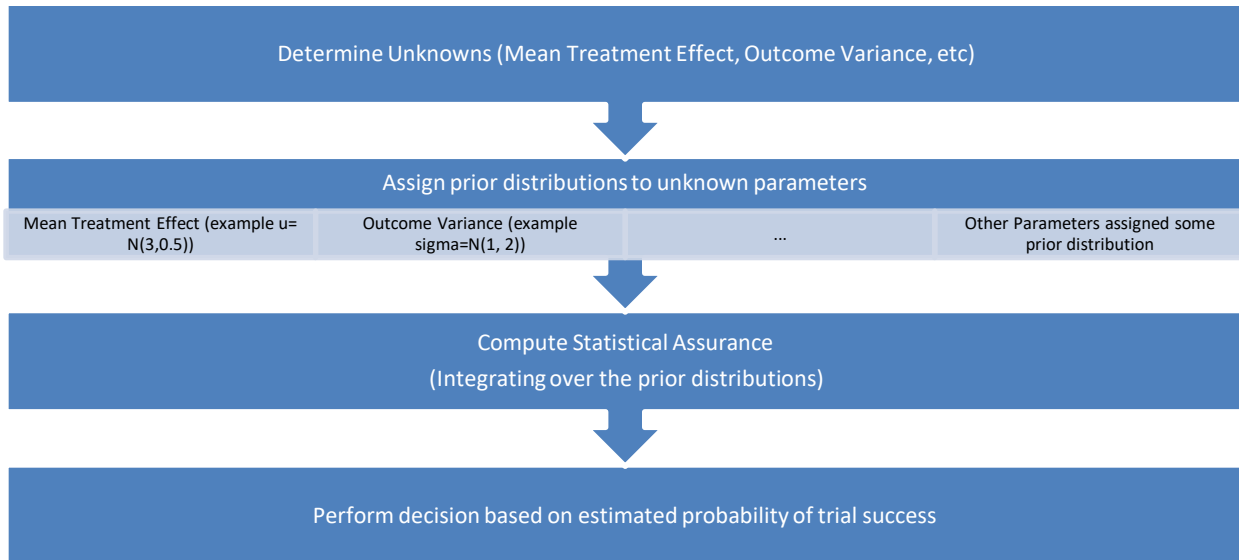
Figure 4. Workflow for Statistical Assurance

In this paper, we introduce the concept of statistical assurance, as described by O'Hagan et al. (2005), which yields a non-conditional probability of a trial's success and we provide some basic implementation in SAS. Statistical assurance is a non-conditional probability that uses a slightly Bayesian framework to replace prior treatment effect and variability estimates (i.e. guessed numbers for the sake of power calculation) with distributions that incorporate all available evidence during the trial planning stages. In other words, instead of assuming a treatment effect of 5%, one can use a Normal(5, 2) distribution to quantify prior beliefs on the expected treatment effect. In this way, available pilot data or expert clinical advice can be better utilized and ultimately run more efficient trials with fewer patients, and save underpowered trials from being run needlessly with too few patients due to misleading power assumptions. One way to think of statistical assurance is that it's the expected power with respect to the prior probability distributions of the assumed parameters. i.e. Why assume of a single treatment effect of 10% when you can "average out" all of the power values across all possible treatment effect values, weighted by their prior probabilities?

For demonstrative purposes, we will work with continuous and binary data, though extensions to time-to-event data do exist (Ren et al., 2013).

Statistical assurance uses Bayesian ideology in a frequentist framework to produce unconditional probabilities of a successful trial, compared to the conditional traditional power calculations that rely on estimates of unknown quantities such as population variances and treatment effects.

ASSURANCE WITH NORMAL DATA AND KNOWN VARIANCES

Consider this example from O'Hagan (2005) of a Phase II superiority trial to assess the effect of a new compound intended to reduce C-reactive protein (CRP) in patients with rheumatoid arthritis. CRP is a measure for disease severity. The outcome variable is CRP reduction after 4 weeks relative to baseline and the main analysis is a two-sided test of superiority at the 5% alpha level.

The population variances for CRP reduction in each of two treatment groups (treatment and placebo) is assumed to known, with values $\sigma_1^2 = \sigma_2^2 = 0.0625$. The protocol specifies that the test should have a power of 80% to detect a treatment effect (reduction in CRP) of 0.2. Traditionally, this leads to a calculated sample size of 25 patients in each group ($n_1 = n_2 = 25$).

To calculate the assurance of this trial (Chuang-Stein, 2006), we can collect prior information about the unknown treatment effect from experts who suggest that the mean treatment effect is expected to be 0.2,

with a variance of 0.0625. If a normal distribution is assumed for the distribution of the treatment effect, then:

$$Assurance = P(Trial\ Success)$$

$$Assurance = P(pvalue < 0.05)$$

Which by law of total probability can be re-written as:

$$Assurance = \int_{-\infty}^{+\infty} P(pvalue < 0.05|\Delta) * P(\Delta|d)d\Delta$$

$$or\ Assurance = \int_{-\infty}^{+\infty} P(difference\ is\ significant|true\ difference)$$
$$* P(true\ difference|observed\ difference)d(true\ difference)$$

Where the sample mean is distributed as:

$$\hat{\delta} = \bar{x}_1 - \bar{x}_2 \sim N(\delta, \tau^2)$$

$$\tau = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Prior distribution for the effect difference (conjugate prior containing prior beliefs):

$$\delta \sim N(m, v)$$

The unconditional distribution (posterior distribution) is then:

$$\hat{\delta} = \bar{x}_1 - \bar{x}_2 \sim N(m, \tau^2 + v)$$

The derivation shown above is a typical example of Bayesian updating with a conjugate normal prior distribution.

The code presented is designed to take input prior beliefs on the treatment difference mean and variance, as well as the within-group variance (in this example the group sample sizes are the same). It is also possible to modify this code to allow for different group sizes using the formulas above.

```
/*pmean: prior mean of the treatment effect (m)
  priorsd: sd of the prior distribution of the effect difference (v)
  popsd: population SD (assumed the same for each treatment
group)(sigma1^2 and sigma2^2)
  posize: group size (n1 and n2)
*/
%macro bprob(pmean,priorsd,posize, popsd);
data _null_;
proc iml;
postsdm=SQRT(2/&posize)*&popsd;   /*calculate posterior standard
deviation of sample treatment difference*/
/*set up and perform numerical integration for assurance*/
start Integrand(delta) global(postsdm);
     postsdm=SQRT(2/&posize)*&popsd;
     return ((1-probnorm((1.96*postsdm - delta)/(postsdm))) *
pdf("normal",delta,&pmean,&priorsd));
finish;
a=.M;
b=.P;
call quad(Assurance, "Integrand", a||b);
```

```
print Assurance;
%mend;

/*test*/
%bprob(0.2,0.25,25,0.25)
```

SAS Code Fragment 1. Assurance of Normal Data with Known Variance

Running the macro with the parameters for the original problem, we compute an assurance of 0.593. In other words, with a group size of 25 the probability of the trial producing positive results is 0.593.
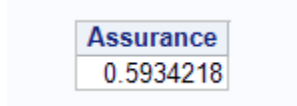
Assurance
0.5934218

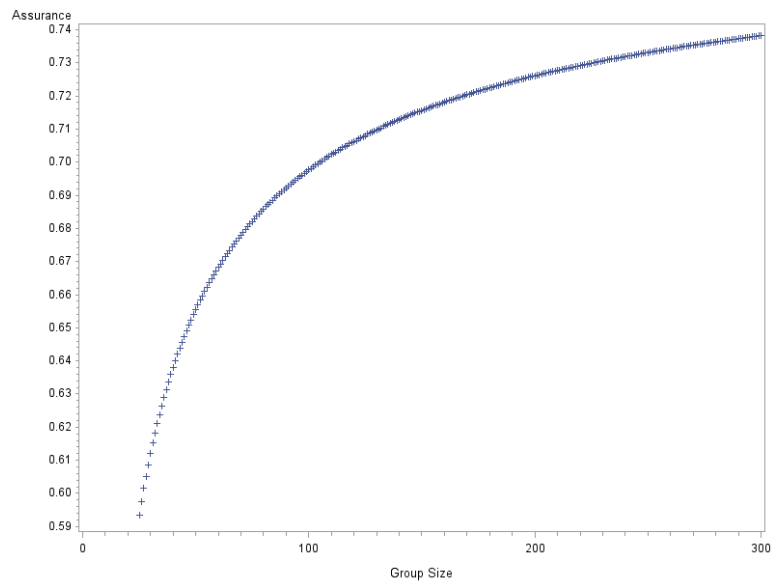**Figure 5. Assurance for a group size of 25**

Figure 6. Assurance plot by group size

Modifying this macro to run through a loop of group sizes for the test run, we can plot assurance by group size as shown above. We notice a diminishing return effect as the curve tends asymptotically towards a maximum possible assurance (i.e. maximum probability of success).

ASSURANCE WITH NORMAL DATA AND UNKNOWN VARIANCES

If the variances are unknown, a different approach must be taken. If equal variances are assumed between treatment groups, we may use the following test statistic:

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\hat{\sigma}^2 = (n_1 + n_2 - 2)^{-1} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}_i\right)^2$$

5

1- Set counter $I = 0$ and select a number of simulations (such as 10 000 000)

2- Sample δ and σ^2 from their joint prior distribution. (i.e. sample them simultaneously)

3- Sample $\bar{x}_2 - \bar{x}_1 \sim N\left(\delta, \left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)\sigma^2\right)$ and $\frac{(n_1+n_2-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$

4- The average of the resulting 10 000 000 statistical powers is the assurance estimate

```sas
%let nsim=1000000;
%let pmean=2.5;
%let psd=7.14;
%let psize=25;
%let posize=128;

data uniform;
do i=1 to &nsim;

/*sample chisq for sigma, since (n-1)*s^2/sigma^2 ~ chisq(n-1)*/
sd=sqrt(((&psize-1)*(&psd*&psd))/rand('CHISQUARE',&psize-1));

/*calculate the standard deviation for the mean*/

psdm=sqrt(2/&psize)*sd; /*priod sd for the mean*/
posdm=sqrt(2/&posize)*sd; /*posterior sd for the mean*/

/*sample the prior*/

Delta=rand('Normal',&pmean,psdm);

/*with the sampled prior, calculate the power*/

power=1-probnorm(((1.96*posdm)-Delta)/posdm);

output;
end;
run;

proc means data=uniform;
    var power;
    output out=ass1 mean=samplemean;
  run;
```

SAS Code Fragment 3. Assurance of Normal Data with Unknown Variance

ASSURANCE FOR BINARY DATA

For binary outcome data, the MCSB (Monte Carlo Simulation Bayesian) algorithm is:

1- Set counter $I = 0$ and select a number of simulations (such as 10 000 000)

2- Sample p_1 and p_2 from their prior distributions (typically beta distributions)

3- Calculate the Z-statistic and the associated statistical power

4- The average of the resulting 10 000 000 statistical powers is the assurance estimate

```sas
%let nsim=1000000; /*number of simulations*/
%let alpha=0.05; /*alpha level*/
%let posize1=200; /*group 1 size)*/
%let posize2=400; /*group2 size*/

data uniform;
do i=1 to &nsim;

/*sample beta for p1*/
p1=rand("Beta",5,20);

/*sample Beta for p2*/
p2=rand("Beta",2,23);

/*calculate z-value*/
zval=(p2-p1)/sqrt((p1*(1-p1)/&posize1)+(p2*(1-p2)/&posize2));

/*calculate the power from the sampled prior*/
quant=quantile("normal",1-(&alpha/2));
power=1-probnorm(-quant+zval);

output;
end;
run;

proc means data=uniform;
    var power;
    output out=ass2 mean=samplemean;
run;
```

SAS Code Fragment 4. Assurance of Binary Data

CONCLUSION

This paper describes the basic clinical scenarios in which statistical assurance can be applied along with their respective SAS implementations and frequentist power calculation comparisons. Namely, this paper covered normal data where variance is known, where variance is unknown, and binary data. Time-to-event data is also possible and will be covered in future papers. Further work must be done to promote and facilitate these methods among practitioners who would benefit from the improved efficiency in trial design and resource management, especially in large multi-phase trials where go/no-go decisions are vital to portfolio success.

REFERENCES

Chuang-Stein, C. 2006. "Sample size and the probability of a successful trial." *Pharmaceut. Statist.*, 5: 305–309. doi:10.1002/pst.232

Chen, D., Ho, S. 2016. "From statistical power to statistical assurance: It's time for a paradigm change in clinical trial design." *Communications in Statistics - Simulation and Computation*. DOI: 10.1080/03610918.2016.1259476

Chen, D., Fraser, M., Cuddeback, G. 2018. "Assurance in Intervention Research: A Bayesian Perspective on Statistical Power." *Journal of the Society for Social Work and Research* 2018 9:1, 159-173

O'Hagan, A., Stevens, J. W. and Campbell, M. J. 2005. "Assurance in clinical trial design." *Pharmaceut. Statist.*, 4: 187–201. DOI:10.1002/pst.175

Ren, S., Oakley, J. 2013. "Assurance calculations for planning clinical trials with time-to-event outcomes", *Statistics in Medicine*. DOI: 10.1002/sim.5916

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Jonathan L. Moscovici
IQVIA
E-mail: jonathan.moscovici@quintiles.com

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