

## Conservative Statistical Power Analysis (SPA) Algorithm: *An Accurate “1-Step” Method with Rules-of-Thumb and Error Analysis Implications*

Alvah C. Bittner, PhD, CPE  
Bittner & Associates  
Kent, WA 98042-3532, USA

Corresponding author's Email: [drbittner@comcast.net](mailto:drbittner@comcast.net)

**Author Note:** Dr. Bittner is a *researcher-practitioner* with a broad-spectrum of interests in methodologies for design and evaluation of workplaces, products, equipment, and systems. His continuing focus – across this spectrum – has been on a) Structuring previously-developed salient information; and b) Efficient research design that addresses significant unanswered questions. Alvah is an *ISOES Past-President and Conference Chair* as well as *Human Factors & Ergonomics Society Fellow*.

**Abstract** *Statistical Power Analyses (SPAs)* estimate required study sample-sizes ( $N$ s) to ensure a high likelihood ( $1-\beta$ , aka. “Power”) of correctly rejecting a no-difference (null) hypotheses with a low target significance level  $\alpha$ . SPAs are often initially conducted to assess relative requirements for success from amongst alternative experimental design, method and/or analysis combinations. SPAs also typically are mandated by Institutional Review Boards (IRBs) as part of ethical ‘risk-benefit considerations. Paralleling SPA concerns, IRB risk-benefit considerations – encountered in ergonomics and safety studies – include ensuring: 1) Minimized numbers of participants put at risk (as also their magnitude); but (2) Participant numbers are also sufficient for a high likelihood of capturing meaningful results. Toward addressing such issues, we outline a new-algorithm that – in contrast to previous rules requiring a secondary accuracy check – yields 1-Step *conservative estimates of required sample sizes (Ns)*. Key to this development is an estimate for Student’s  $t$ -distribution – with  $dF$  degrees-of-freedom – in terms the classical Normal-Gaussian ( $Z$ )-Distribution:  $t(\alpha, dF) \approx (dF/(dF-3))^{1/2} Z(\alpha)$ . The utility of this approximation is evaluated against exact corresponding  $t$ -distribution “threshold” values:  $dF = 10$  to  $\infty$  for salient  $\alpha, 1-\beta$ -combinations. Combinations include 1- and 2-tailed for both 0.05, 0.80 and .01, 0.90 (the latter is required minimum for medical and low-replication-probability-studies). Subsequently, taking advantage of the  $t(\alpha, dF)$  approximation, required  $N$  may be expressed in terms of an equation with functions of  $N$  on both sides. This is readily resolved by (1) Isolation of  $N$  in terms of a ratio on one side of the expression and (b) Synthetic division that serves to yield the desired conservative 1-Step algorithm. The 1-Step algorithm is explored following its derivation. This includes updating to classical 1- and 2-tailed ( $\alpha=0.05$ ,  $1-\beta=0.80$ ) “rules-of-thumb; as well as, new 2-tailed  $\alpha=0.01$ ,  $1-\beta=0.90$  variants for medical and unlikely-to-be-replicated research. Also explored are error-analysis implications of the algorithm when input terms are estimated from previous research.

Recommended are adoptions of both: (1) the newly derived conservative 1-Step Rules-of-Thumb and (2) associated error-analysis considerations.

**Keywords:** Statistical Power Analyses (SPA), Conservative Algorithm, New Rules of Thumb

## 1. INTRODUCTION

### 1.1 Statistical Power Analysis

*Statistical Power Analysis (SPA)* estimates the smallest sample-size ( $N$ ) for a study given: a required significance level ( $\alpha$ ), effect size ( $\Delta$ ) and a high likelihood “power ( $1-\beta$ )” of rejecting null hypotheses (e.g., Scheffe’, 1959; Cohen, 1988; 1992). Toward minimizing *research costs* (financial and time), SPAs may be conducted to access sample-size requirements from amongst alternative experimental design, method and/or analysis options. In this regard, we have previously explored a wide range of options for *more powerful* human-factors/ergonomics (HF/E) research (e.g. Bittner. Bramwell et al., 1998;

Bittner, Winn et al., 2003; 2004 a&b). Relatedly, SPAs are typically mandated by *Institutional Review Boards* (IRBs) as part of ethical ‘participant *risk-benefit considerations* (e.g., Vollmer & Howard, 2010). In this regard, risk-benefit considerations – encountered in ergonomics and safety studies – include ensuring: (1) Minimized numbers of participants put at risk (as also their magnitude); but (2) Participant numbers are also sufficient for providing a high likelihood (power) of capturing meaningful results. These – together with research cost minimizations – have encouraged the development of technically accurate methods (Scheffe’, 1959; Cohen, 1988;1992; Wang, & Ji, 2020).) as well as quick, but less trustworthy “approximations” (Lehr, 1992; Dunlap & Kennedy, 1995; Bittner & Bittner, 2009). SPA Rules-of-Thumb (ROTs) – and kindred approximate SPA tools – long have been suggested for *initially* “sizing” an experimental design (*1<sup>st</sup> Step*), but with a *following-up* (*2<sup>nd</sup> Step*) with a technically accurate method (e.g., Lehr, 1992; Van Belle, 2011).

## 1.2 Purpose

We have three primary goals in the following. Our first (Sec 2.1) is - in concert with demonstrating the approximation:  $t(\alpha, dF) \approx (dF/dF-3)^{1/2}Z(\alpha)$  – to derive the new 1-Step algorithm. Our second (Sec 2.2) is to consider immediate implications of this algorithm for classical ROTs as well as 1- and 2-tailed variants (per, Pocock, 1988) for “...medical and unlikely-to-be-replicated research” (i.e.,  $\alpha=0.01$ ,  $1-\beta=0.90$ ). Our final goal (Sec. 3) is to 1) Recommended near-term adoption of our newly derived conservative 1-Step rules-of-thumb, and 2) Longer-term explorations of SPA error-analyses that prospectively could more robustly estimate required sample sizes.

## 2. 1-STEP DERIVATION AND IMPLICATIONS

### 2.1 Derivation 1-Step Algorithm

Our generalized derivation will be conducted in three steps. In Sec 2.1.1, we will initiate our derivation by consideration of a “clumsy” extension (Eq. 1c) of a classical case (i.e., Eqs. 1a&b). This clumsiness arises as Eq. 1c has functions of required N on both its sides; whereas, Eqs. 1a&b escape this with the assumption of an effectively asymptotic ( $\rightarrow \infty$ ) required sample size (N). In Sec 2.1.2, we explore the effectiveness of an approximation use to resolve the clumsiness of (Eq. 1c). This sets the stage for final derivation of our 1-Step algorithm in 2.1.3 and subsequent exploration of its implications in Section 2.2.

#### 2.1.1 First Steps.

The derivation – of required group sample size N – begins with a readily understood U-statistic outlined below [Eq. 1a], that has been widely shown to apply for comparison of the means of two independent groups where the sample-size is treated as if effectively asymptotically large (e.g., Bittner & Bittner, 2009).

$$U = \frac{\Delta}{[2\sigma^2/(N)]^{1/2}} = Z(\alpha) + Z(1-\beta) \quad (\text{Eq. 1a})$$

Where:  $\Delta$  = a minimum “critical difference” identified by practitioner-researchers as important to detect [e.g. difference between two population means ( $\mu_1-\mu_2$ )].

$2\sigma^2/N$  = *exact* variance of the critical difference  $\Delta$  [e.g., for difference of two independent means – with common variance  $\sigma^2$  and N observations each condition], and

$Z(\alpha)$  is the Gaussian z-score value appropriate for  $\alpha$  significance level (e.g., 0.05) given number of tails (i.e., 1.645 1-tailed or 1.960 2-tailed in 0.05 case), and

$Z(1-\beta)$  is the Gaussian z-value for desired power level  $1-\beta$  (e.g., 0.80 or 0.90).

Here, one might note that the left side of Eq 1a is set up much like a classical U-test for significance, but with a threshold value of  $Z(\alpha) + Z(1-\beta)$  on the right (vs.  $Z(\alpha)$ ). The requirement for adding  $Z(1-\beta)$  – to achieve a  $1-\beta = 0.80$  – may be appreciated by consideration of Figure 1. Specifically, the 1.96 cutoff ( $\alpha=0.05$ , 2-tailed) with no difference (I), may be seen to require displacement (0.84) further to 2.80 (1.96+0.84) to assure that the percentage of its cases falling below 1.96 is 20% (i.e.,  $\beta=0.20$ ) and power  $1-\beta=0.80$ ). With this in mind, we note that one may rearrange terms of Eq. 1a to solve for N:

$$N = 2[Z(\alpha) + Z(1-\beta)]^2(\sigma^2/\Delta^2)$$

(Eq. 1b)

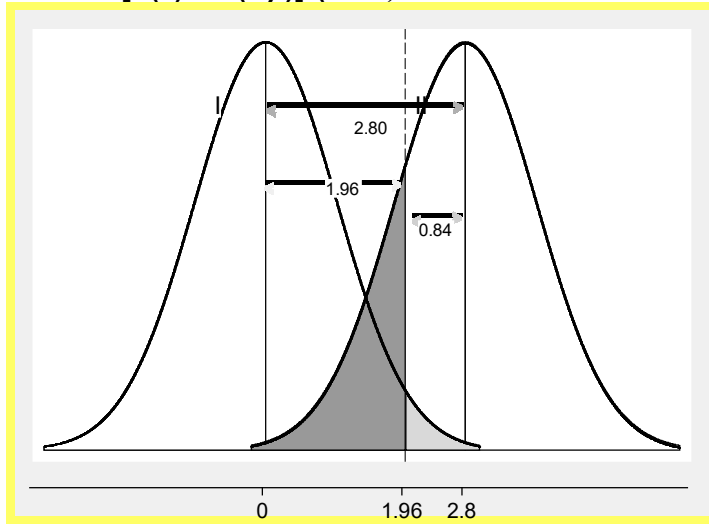


Fig. 1. Illustration of 2-tailed  $\alpha=0.05, 1-\beta=0.80$  ( $Z(\alpha)+Z(1-\beta)=2.8$ )

Eq. 1b – as we will see later – leads to a number of ROTs that have previously proven useful; albeit, with recommendations for 2<sup>nd</sup> steps (Lehr, 1992; Van Belle, 2011). Accuracy issues occur because  $N$  is not effectively asymptotic which is addressed by substituting the more general  $t(dF, \alpha)$  and  $t(dF, 1-\beta)$  for  $Z(\alpha)$  and  $Z(1-\beta)$  in Eq. 1b to obtain:

$$N = 2[t(dF, \alpha) + t(dF, 1-\beta)]^2(\sigma^2/\Delta^2)$$

(Eq. 1c)

where  $dF = N-1$  is the most conservative estimated vis-à-vis the *Brehens-Fischer Problem* (e.g., Welch, 1938). The complication “clumsiness” with this form (Eq 1c) is that the  $t$ -values are a function of  $N$  (i.e.,  $N$  represented on both sides). In the next section (2.1.2) we will consider our robust  $t$ -distribution approximation that will prove helpful in resolving this issue.

### 2.1.2 $t$ -Distribution Approximation.

Table 1 below explores the accuracy of a  $t$ -distribution approximation selected to resolve the noted clumsiness (i.e.,  $N$  represented on both sides of Eq 1c). Suggested by the form of an earlier approximation (Bittner & Bittner, 2009), this is:

$$t(\alpha, dF) \approx (dF/(dF-3))^{1/2}Z(\alpha)$$

(Eq. 1d)

Table 1. **Exact** and **Approximate** Significance ( $\alpha$ ) and Power-Shift ( $1-\beta$ )  $t$ -Values

DEGREES OF FREEDOM ( $dF$ )	ONE-TAILED $t(\alpha, dF)$		TWO-TAILED $t(\alpha, dF)$		POWER $t(1-\beta, dF)$	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$1-\beta = 0.8$	$1-\beta = 0.9$
Infinite	1.645 <i>1.645</i>	2.326 <i>2.326</i>	1.960 <i>1.960</i>	2.576 <i>2.576</i>	0.841 <i>0.841</i>	1.282 <i>1.282</i>
100	1.660 <i>1.670</i>	2.364 <i>2.362</i>	1.984 <i>1.990</i>	2.626 <i>2.615</i>	0.845 <i>0.854</i>	1.290 <i>1.301</i>
25	1.708 <i>1.745</i>	2.485 <i>2.480</i>	2.060 <i>2.089</i>	2.787 <i>2.746</i>	0.856 <i>0.897</i>	1.316 <i>1.367</i>
15	1.753	2.602	2.131	2.947	0.866	1.341

	1.839	2.601	2.191	2.880	0.940	1.433
10	1.812	2.764	2.228	3.169	0.879	1.372
	1.966	2.781	2.342	3.079	1.005	1.532

Examining Table 1, one may note that all values for the **approximation** (in blue) are near those **exact** ( $dF = 10$  to  $\infty$ ). Most importantly, where ( $\alpha=0.01$ ) approximations are less than exact, the approximate  $t(1-\beta, dF)$  exceeds the exact sufficiently such that substituting the approximation for term  $[t(dF, \alpha) + t(dF, 1-\beta)]$  in Eq 1c will result in a conservative estimate for  $N$ . We make this substitution – and complete our basic derivation --in the next section (2.1.3).

### 2.1.3 Conservative Algorithm Derivation

Derivation of our conservative algorithm may now proceed with substitution of the t-distribution approximation (i.e., Eq. d) for the exact values in Eq. 1c to initially obtain:

$$N \approx 2[(dF/(dF-3))^{1/2}Z(\alpha) + (dF/(dF-3))^{1/2}Z(1-\beta)]^2(\sigma^2/\Delta^2) \quad (\text{Eq. 1e})$$

where it may be recalled that in the context of Eq. 1c, (i.e. Welch, 1938) that conservatively  $dF = N-1$ . Making this latter substitution, and collecting terms, one obtains:  $N \approx (N-1)/(N-4) [2[Z(\alpha) + Z(1-\beta)]^2(\sigma^2/\Delta^2)]$  or equivalently

$$N[(N-4)/(N-1)] \approx 2[Z(\alpha) + Z(1-\beta)]^2(\sigma^2/\Delta^2) \quad (\text{Eq. 1f})$$

where the right-hand expression is that seen with Eq. 1b, upon which traditional ROTs are based! Further, noting via synthetic division that the left-hand part of the expression

$$N[(N-4)/(N-1)] = N-3 - (3/N-1) \leq N-4 \text{ for all } N \geq 4$$

Hence, simply adding 4 to the traditional asymptotically based estimates – seen earlier (Eq. 1b) -- will quite generally provide a conservative estimate (Eq. 2) for required  $N$ , within certainly the range  $N \geq 10$  where the approximation (Eq. 1d) was shown conservative.

$$N = \{2[Z(\alpha) + Z(1-\beta)]^2(\sigma^2/\Delta^2)\} + 4 \quad (\text{Eq. 2})$$

This [Eq. 2] is the general form which will be explored regarding implications in the following Section (2.2)

## 2.2 Implications

Drawn in the following are two implications of Eq 2. First, in 2.2.1, are implications regarding our conservative 1-Step exact and revised rules-of-thumb estimates for required sample sizes ( $N$ ). The second, in 2.2.2, considers an obvious – but relatively unexplored – utilization of the rules-of-thumb for error-analyses toward addressing specific component (re:  $\sigma^2$  and/ or  $\Delta$ ) variability – issues that frequently occur in HF/E.

### 2.2.1 1 Step and Classical Sample Required Sample Sizes Computations

Table 2 summarizes – for four cases of special interest – the computations for the Coefficients, Asymptotic sample-size and our conservative  $N$  Requirements. Cases include combinations of one- and two-tail tests with most typically applied ( $\alpha = 0.05$ ,  $1-\beta = 0.8$ ) and that ( $\alpha = 0.01$ ,  $1-\beta = 0.9$ ) which is typically required for medical and low-replication-probability-studies. Values for the  $Z(\alpha)$  and  $Z(1-\beta)$  in computing the coefficient of Eq. 2, were conveniently obtained from the first row of Table 1 as  $Z(\alpha) = t(\alpha, dF=\infty)$ . For 2-tailed ( $\alpha=0.05$ ,  $1-\beta=0.8$ ), we can see from the table that  $(Z(\alpha) + Z(1-\beta)) = 1.960 + 0.841 = 2.80$  (rounded) which is the case illustrated earlier in Figure 1. The coefficient of Eq. 2 is easily computed as twice the square of 2.801 or 15.7 ( $\approx 16$ ) as shown under the coefficient column for 2-tailed with  $\alpha=0.05$ ,  $1-\beta=0.8$ .

Table 2, under “Asymptotic  $N$  Requirement”, one may note two approximations that (while requiring a second step) have long been broadly employed (Lehr, 1992; Van Belle, 2011). Also delineated are two cases ( $\alpha = 0.01$ ,  $1-\beta = 0.9$ ),

that we believe are original here. Our Conservative 1-Step N Requirements are summarized in the last column where only rounded values are shown for the  $\alpha=0.01$  as remarkably close to the exact.

Table 2. Classical and 1-Step Required Sample-Size (N) Computations and Associated Rules-of-Thumb.

	COEFFICIENT $2[Z(\alpha) + Z(1-\beta)]^2$	ASYMPTOTIC N REQUIREMENT	CONSERVATIVE 1-STEP N REQUIREMENT
<b>ONE-TAILED</b>			
$\alpha = 0.05, 1-\beta = 0.8$	12.36 $\approx 12.5$	$N = 12.36(\sigma^2/\Delta^2)$ $\approx 12.5 (\sigma^2/\Delta^2)^*$	$N = 12.36(\sigma^2/\Delta^2)+4$ $\approx 12.5 (\sigma^2/\Delta^2)+4^{**}$
$\alpha = 0.01, 1-\beta = 0.9$	17.3 $\approx 17$	$N = 17.13(\sigma^2/\Delta^2)$ $\approx 17 (\sigma^2/\Delta^2)$	$N \approx 17(\sigma^2/\Delta^2) + 4^{**}$
<b>TWO TAILED</b>			
$\alpha = 0.05, 1-\beta = 0.8$	15.7 $\approx 16$	$N = 15.7(\sigma^2/\Delta^2)$ $\approx 16 (\sigma^2/\Delta^2)^*$	$N = 15.7(\sigma^2/\Delta^2) + 4$ $\approx 16(\sigma^2/\Delta^2) + 4^{**}$
$\alpha = 0.01, 1-\beta = 0.9$	21.0 $\approx 21$	$N = 21 (\sigma^2/\Delta^2)$	$N = 21(\sigma^2/\Delta^2) + 4^{**}$

\* Classical Rule of Thumb (Lehr, 1992; Van Belle, 2011). \*\*1-Step Rule of Thumb

### 2.2.2 Error-Analysis

Introduction of error analysis – into considerations of required sample sizes (N) – may seem unusual until realizations that  $\sigma^2$  and/or  $\Delta^2$  are frequently only estimates. In our HF/E experience,  $\sigma^2$  is most often estimated from an earlier study (e.g., performance or risk); whereas,  $\Delta^2$  may be a rough value drawn from other study types (e.g., marketing). Error-analysis, in these situations as we will see, suggests why research oft falls short of power (e.g. Schneck, 2023). More importantly, it also can be useful in making more accurate N sizing selections. Applying (to Eq. 2 after subtracting 4 from both sides) the classical approach (Deming, 1943, pp. 37-48; Cameron, 1982, esp. p.549), we find the total differential form:

$$d(N-4)/(N-4) = dC/C + d(\sigma^2)/\sigma^2 - d(\Delta^2)/\Delta^2 \quad (\text{Eq. 3})$$

Where  $C = 2[Z(\alpha) + Z(1-\beta)]^2$  is a constant (hence  $dC/C = 0$ ). Continuing our general approach: (i) Noting  $d(N-4) = d(N)$ , (ii) Replacing differentials with small incremental values, (iii) Squaring both sides, (iv) Taking expected values, and (v) Assuming zero covariance between estimated  $\sigma^2$  and  $\Delta^2$ , one may obtain:

$$\text{Var}(N) \approx ((N-4)/\sigma^2)^2 \text{Var}(C) + ((N-4)/\Delta^2)^2 \text{Var}(\Delta^2) \quad (\text{Eq. 4})$$

Where the variances  $\text{Var}(\sigma^2)$  and/or  $\text{Var}(\Delta^2)$  can come from previous studies that produced the input  $\sigma^2$  and  $\Delta^2$  values. The implications of Eq. 4 may be considered in an exemplar 2-tailed  $\alpha = 0.05, 1-\beta = 0.80$  case, where desired  $\Delta^2 = 0.64$  for a Jackknife estimated  $\sigma^2 = 1$  (here normalized for confidentiality). Employing our new Table 2 1-Step rule,  $N = 16 (\sigma^2/\Delta^2) + 4 = 29$ . However, in light a modest associated Jackknife estimated  $\text{Var}(\sigma^2) = 0.01$ , Eq. 4 would imply  $\text{Var}(N) \approx 29^2(0.01) \approx 6.25$  which would correspond to Standard Deviation of the Required N = 2.5. This – in addition to maybe prompting a search for a more accurate estimate of  $\sigma^2$  – would arguably support a significantly increased required sample size (e.g., ~31). We and our colleagues – not surprisingly – have found kindred error-analyses “enlightening.”

## 3. CONCLUSION AND RECOMMENDATIONS

We met our goals as outlined earlier. First – after demonstrating of the functional accuracy of  $t(\alpha, dF) \approx (dF/(dF-3))^{1/2}Z(\alpha)$  in Table 1 – we completed the derivation of the 1-step algorithm (Eq. 2). Second, using this (Eq. 2), we first employed it in developing new 1-Step Rules-of-Thumb (Table 2). Of note, these new ROTs included novel 1- and 2-tailed variants applicable for medical and unlikely-to-be-replicated research (i.e.,  $\alpha=0.01, 1-\beta=0.90$ ). Further, building off Eq. 2, a classical “error-analysis” approach was used to derive Eq. 4, which is applicable for exploring  $\text{Var}(N)$  as a function of the variability of SPA input values (i.e.,  $\sigma^2$  and  $\Delta^2$ ). Toward illustrating applications of Eq. 4, an exemplar was offered that

pointed toward a requirement to increase initially computed required  $N_s$ , to account for inherent variances in the input values (e.g. when drawn from previous studies). It is clear – in meeting goals – that our results arguably offer opportunities for enhancing the processes of SPAs. ***We consequently recommend adoption of both (1) Our newly derived conservative 1-step Rules-of-Thumb, and 2) Associated error-analysis considerations.***

#### 4. REFERENCES.

- Bittner A.C. & Bittner R.C.L. (2009). “Right-sizing” research studies: Assuring adequate but not grossly-overlarge sample-sizes. *Proceedings 53<sup>rd</sup> Annual Meeting Human Factors and Ergonomics*. Santa Monica, CA: HFE
- Bittner, A.C., Bramwell, A.T., Morrissey, S.J. & Winn, F.J. (1998). Options for more powerful human-factors/ ergonomics independent group studies. In S. Kumar (Ed.), *Advances in occupational ergonomics and safety 2* (pp. 3-10). Amsterdam: IOS Press.
- Bittner, A.C., Winn, F.J. & Morrissey, S.J. (2003). Proportionally reducing sample-size requirements by increasing dependent variable reliability. *Proc 47th Ann. Meet Human Factors and Ergo. Soc.* (CD-ROM, pp. 2000-2004). Santa Monica, CA: HFES
- Bittner, A.C., Winn, F.J. & Morrissey, S.J. (2004a). Options for more powerful human-factors/ ergonomics independent-groups studies: Increasing dependent variable reliability. *Ergonomia*, 26(2). 141-149.
- Bittner, A.C., Winn, F.J., Bittner, R.C.L. & Lundy, N.C. (2004b). Theory-Based directional-testing: Revisiting a primary option for efficient human-factors/ergonomics (HF/E) studies. *Proc 48th Ann. Meet. Human Factors and Ergo Society (CD-ROM)*. Santa Monica, CA: HFES
- Cameron, J.M. (1982). Error analysis. In: *Encyclopedia of Statistical Sciences*, Vol 2 (pp.545-551). New York, NY: John Wiley & Sons. .
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1992). A power primer. *Psych. Bull.*, 112(1), 155-159.
- Deming, W.E. (1943). *Statistical Adjustment of Data*. New York, NY: Dover Publications
- Dunlap, W. & Kennedy, R.S. (1995). Testing for statistical power. *Ergonomics in Design*, 3-6ff..
- Lehr, R. (1992). Sixteen S-squared over D-squared: A relation for crude sample size estimates. *Stat. in Med.*, 11(8), 1099-1102.
- Pocock, S.J. (1988). *Clinical Trials: A Practical Approach*. New York, NY: Wiley & Sons.
- Scheffe’ H. (1959) *The Analysis of Variance*. New York, NY: John Wiley & Sons)
- Schneck, A. (2023). Are most publications false? Trends in statistical power, publication selection bias, and the false discovery rate in psychology (1975-2017). *PLOS ONE*, 18(10): e0292717. [Doi:10.1371/journal.pone.0292717](https://doi.org/10.1371/journal.pone.0292717)
- Van Belle, G. (2011). *Statistical rules of thumb*. New York, NY: John Wiley & Sons. [ISBN 978-0-470-37796-3](https://doi.org/10.1002/9780470377963).
- Vollmer, S.H. & Howard, G (2010). Statistical Power, the Belmont Report, and the Ethics of Clinical Trials. *Sci Eng Ethics* 16, 675–691. <https://doi.org/10.1007/s11948-010-9244-0>
- Wang, X., & Ji, X. (2020). Sample size estimation in clinical research: from randomized controlled trials to observational studies. *Chest*, 158(1), S12-S20.
- Welch, B.L. (1938). The significance of the difference of two means when the population variances are unequal. *Biometrika*, 29(3/4), 350-362.