

Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach

KEN KELLEY

Indiana University, Bloomington, Indiana

The accuracy in parameter estimation approach to sample size planning is developed for the coefficient of variation, where the goal of the method is to obtain an accurate parameter estimate by achieving a sufficiently narrow confidence interval. The first method allows researchers to plan sample size so that the expected width of the confidence interval for the population coefficient of variation is sufficiently narrow. A modification allows a desired degree of assurance to be incorporated into the method, so that the obtained confidence interval will be sufficiently narrow with some specified probability (e.g., 85% assurance that the 95% confidence interval width will be no wider than ω units). Tables of necessary sample size are provided for a variety of scenarios that may help researchers planning a study where the coefficient of variation is of interest plan an appropriate sample size in order to have a sufficiently narrow confidence interval, optionally with some specified assurance of the confidence interval being sufficiently narrow. Freely available computer routines have been developed that allow researchers to easily implement all of the methods discussed in the article.

The coefficient of variation is a dimensionless number that quantifies the degree of variability relative to the mean. The population coefficient of variation is defined as

$$\kappa = \frac{\sigma}{\mu}, \quad (1)$$

where σ is the population standard deviation and μ is the population mean. The typical sample estimate of κ is given as

$$k = \frac{s}{M}, \quad (2)$$

where s is the sample standard deviation, the square root of the unbiased estimate of the variance, and M is the sample mean. Equations 1 and 2 are sometimes multiplied by 100 so that the ratio of the standard deviation to the mean is expressed in terms of a percentage.

The coefficient of variation has long been a widely used descriptive and inferential quantity in various areas of the biological and medical sciences. Compared with several other effect size measures, the coefficient of variation has not historically been widely used in behavioral, educational, or social sciences. However, some research questions within the behavioral, educational, and social sciences lend themselves to being addressed with the coefficient of variation. At least in part due to the current emphasis in many parts of the literature on the interaction of biological and psychological systems as explanatory factors of behavior (e.g., Frith & Frith, 2001; Kosslyn et al., 2002; Salmon & Hall, 1997; etc.), use of the coefficient of variation will likely continue to increase.¹

An example of where the coefficient of variation is often used is for assays, which are procedures that measure certain designated properties of biological components. The results of repeated trials of such assays are many times reported in terms of the coefficient of variation, because the standard deviations of assays generally increase (or decrease) proportional to the mean increase (or decrease; Reed, Lynn, & Meade, 2002). Coefficients of variation are generally based on a single measurement from different individuals (i.e., an *interindividual* coefficient of variation). However, *intraindividual* coefficients of variation, where repeated measures for the same individual are obtained, are also possible. One place where an intraindividual coefficient of variation is of interest is before and after a treatment to evaluate the effectiveness of the treatment (Reed et al., 2002). In a psychiatry setting, Volkow et al. (2002) used the coefficient of variation to compare patterns of homogeneity/heterogeneity in brain metabolism for Alzheimer's disease patients with those in a control group without Alzheimer's. It was shown that the coefficient of variation was larger across the entire cortex, but that there were smaller coefficients of variation in temporal and parietal cortices for Alzheimer's disease patients. As more researchers move to a biology-psychology model of behavior, assays will likely become more prevalent in the behavioral sciences.

Of course, the coefficient of variation need not be restricted to biological systems. In a classic experimental psychology setting, Babkoff, Kelly, and Naitoh (2001)

K. Kelley, kkiii@indiana.edu

reported the effects of sleep deprivation on a four-choice reaction time (RT) experiment to assess performance stability over time between a placebo and two treatment groups that received stimulants. The coefficient of variation was used for a subset of the analyses because mean performance level is postulated to be less affected by sleep deprivation than is the trial-to-trial variance (Babkoff et al., 2001; Dinges & Kribbs, 1991). Babkoff et al. went on to model the coefficient of variation (the measure of performance stability) over time as a function of group membership. Similarly, Hayashi (2000) examined the coefficient of variation for choice RT when participants were using benzodiazepine in an effort to manipulate their cognitive states. In an educational setting, Monchar (1981) used the coefficient of variation to quantify educational inequality in the context of political instability in 39 countries at various times from the late 1950s until the early 1970s. The idea was that countries with greater political instability would have more educational inequities (larger coefficients of variation), using the rate of enrollment for reporting regions within the country, than would more-stable countries.² The coefficient of variation is also used for quantifying risk sensitivity, which is especially helpful when comparing diverse populations (Weber, Shafir, & Blais, 2004). With two experiments, Weber et al. (2004, p. 430; see also Shafir, 2000) show that risk sensitivity of human beings, as with other types of animals, becomes strongly proportional to the coefficient of variation when they learn about choice alternatives. Finally, a comprehensive review of the demography and diversity literature shows that the coefficient of variation is one of the most widely used ways of assessing group-based demographic differences (Williams & O'Reilly, 1998). As can be seen, the coefficient of variation has been used in a wide variety of contexts.

Even though the estimated coefficient of variation can be a useful measure, perhaps the greatest use of it as a point estimate (like most point estimates) is to construct a confidence interval for the population quantity.³ A confidence interval provides much more information about the population value of the quantity of interest than does a point estimate (e.g., Cohen, 1994; Hunter & Schmidt, 2004; Kirk, 2001; Meehl, 1997; Schmidt, 1996; Smithson, 2001; Steiger, 2004; Task Force on Reporting of Research Methods in AERA Publications, 2006; Thompson, 2002; Wilkinson & the American Psychological Association Task Force on Statistical Inference, 1999). The more accurate the estimated quantity, the more information is known about the population quantity. Because the goal of most research is to implicitly or explicitly learn as much as possible about the parameter of interest, obtaining an accurate parameter estimate should generally be of utmost concern in applied research.⁴

The purpose of the present article is to offer an approach to sample size planning for the coefficient of variation, where the goal is to obtain a sufficiently narrow confidence interval that illustrates the expected accuracy with which the parameter has been estimated. In the context of parameter estimation, accuracy is defined as the square root of the mean square error, which is a function of both precision and bias. Precision is inversely

related to the variance of the estimator and bias is the systematic discrepancy of an estimator and the parameter it estimates (Rozeboom, 1966). As the confidence interval width decreases, holding constant the confidence interval coverage, the estimate is contained within a narrower set of plausible parameter values and the expected accuracy of the estimate improves (i.e., the root mean square error is reduced). Thus—provided the confidence interval procedure is exact—when the width of the $(1 - \alpha)100\%$ confidence interval decreases, the expected accuracy of the estimate necessarily increases. The goal of this approach is to plan the necessary sample size so that the estimated coefficient of variation accurately reflects the corresponding population value by achieving a sufficiently narrow confidence interval. This approach to sample size estimation has been termed *accuracy in parameter estimation* (AIPE), because when the width of the $(1 - \alpha)100\%$ confidence interval decreases, the expected accuracy of the estimate increases (Kelley & Maxwell, 2003, in press; Kelley, Maxwell, & Rausch, 2003; Kelley & Rausch, 2006).

Probabilistically, holding everything else constant, the narrower a confidence interval, the higher the degree of expected accuracy of the obtained parameter estimate. The use of the term “accuracy” in this context follows the usage of the term given by Neyman (1937) when he discussed the “accuracy of estimates” in his seminal work on the theory of confidence intervals. He stated that “the accuracy of estimation corresponding to a fixed value of $1 - \alpha$ may be measured by the length of the confidence interval” (Neyman, 1937, p. 358; notation changed to reflect current usage). The problem that the present article solves is planning sample size so that the confidence interval for the coefficient of variation is sufficiently narrow, and thus that the obtained estimate has a sufficient degree of expected accuracy. The overarching goal of AIPE is thus to avoid obtaining confidence intervals that are “embarrassingly large” (Cohen, 1994, p. 1002) and of limited usefulness. Note that the general goal of AIPE has nothing to do with rejecting a null hypothesis, which is the stated goal of power analysis (e.g., Cohen, 1988; Kraemer & Thieman, 1987; Lipsey, 1990; Murphy & Myers, 1998). The goal of planning sample size so that the confidence interval is sufficiently narrow (i.e., the AIPE approach) is a fundamentally different process from the power analytic approach—that is, planning sample size so that the confidence interval does not contain the null value.

The first sample size planning method developed determines the necessary sample size so that the expected width of the confidence interval for the coefficient of variation is sufficiently narrow. A modified sample size procedure is then developed so that there is some specified level of assurance (i.e., a probabilistic statement) that the obtained confidence interval will be sufficiently narrow (e.g., 85% assurance that the 95% confidence interval for κ will be no wider than 0.10 units). General methods are developed and algorithms given so that the necessary sample size can be determined in any particular situation. A series of tables for necessary sample size are provided for a variety of scenarios that may help researchers planning studies in

which the coefficient of variation is of interest to achieve a sufficiently narrow confidence interval.⁵

Estimation and Confidence Interval Formation for κ

Sokal and Braumann (1980, p. 51) derive the expected value of k for normally distributed data given κ and N as

$$E[k(\kappa, N)] = \kappa \left[1 - \frac{1}{4(N-1)} + \frac{1}{N} \kappa^2 + \frac{1}{2(N-1)^2} \right] = \tilde{k}_{(\kappa, N)}, \tag{3}$$

where N is the sample size. Equation 3 will become useful momentarily when developing the sample size planning procedure. Although Equation 2 is the estimate typically reported for the coefficient of variation, as can be seen from Equation 3, k is a (negatively) biased quantity. A “nearly unbiased estimate” (Haldane, 1955, p. 484; Sokal & Braumann, 1980) of κ can be obtained with a simple correction to k ,

$$k_u = \left(1 + \frac{1}{4N} \right) k, \tag{4}$$

where k_u is the nearly unbiased estimate of κ . When reporting point estimates of κ , k_u is the recommended quantity. However, as will be discussed, k is used in the formation of confidence intervals for κ .

Although k_u provides a nearly unbiased estimate of κ , a point estimate in and of itself does not convey any information regarding the uncertainty with which the parameter has been estimated. As previously discussed, the use of confidence intervals is especially helpful for conveying information about the uncertainty with which an estimator estimates the population parameter of interest. Although κ will almost never be known exactly, a range of plausible values can be obtained by forming a confidence interval for κ . Until recently, confidence intervals for κ were prohibitively difficult to obtain. However, as a result of recent advances in statistical software, such intervals can be easily found with several software titles that have noncentral t distribution routines. The remainder of this section discusses how a confidence interval can be formed for the coefficient of variation.

It can be shown that the coefficient of variation follows a noncentral t distribution, when the parent population of the scores is normally distributed, with noncentrality parameter

$$\lambda = \frac{\sqrt{N} \mu}{\sigma} = \frac{\sqrt{N}}{\kappa} \tag{5}$$

with ν degrees of freedom, where $\nu = N - 1$ (Johnson & Welch, 1940; McKay, 1932). An estimate of the noncentrality parameter can be obtained by substituting the typical estimate of κ (i.e., k) for the population value in Equation 5:

$$\hat{\lambda} = \frac{\sqrt{N}}{k}, \tag{6}$$

where $\hat{\lambda}$ is the estimated noncentrality parameter. Given the estimated noncentrality parameter, a $(1 - \alpha)100\%$ confidence interval can be formed for the population noncentrality parameter by making use of the *confidence interval transformation principle* and the *inversion confidence interval principle*. These two principles are nicely delineated in Steiger and Fouladi (1997) and in Steiger (2004).⁶

To illustrate the inversion confidence interval principle, let $1 - \alpha$ be the confidence interval coverage with $\alpha_L + \alpha_U = \alpha$, where α_L is the proportion of times that κ will be less than the lower confidence limit and α_U the proportion of times that κ will be greater than the upper confidence limit in the confidence interval procedure. In most cases $\alpha_L = \alpha_U = \alpha/2$, but that need not be the case (e.g., when $\alpha_L = 0$ and $\alpha_U = .05$, or vice versa, for one-sided confidence intervals). The confidence bounds for λ are determined by finding the noncentrality parameter whose $1 - \alpha_L$ quantile is $\hat{\lambda}$ (for the lower bound of the confidence interval) and by finding the noncentrality parameter whose α_U quantile is $\hat{\lambda}$ (for the upper bound of the confidence interval). Stated another way, the lower confidence bound for λ is the noncentrality parameter that leads to $t_{(1-\alpha_L, \nu, \lambda_L)} = \hat{\lambda}$ and the upper confidence bound for λ is the noncentrality parameter that leads to $t_{(\alpha_U, \nu, \lambda_U)} = \hat{\lambda}$, where $t_{(1-\alpha_L, \nu, \lambda_L)}$ is the value of the noncentral t distribution at the $1 - \alpha_L$ quantile with noncentrality parameter λ_L and $t_{(\alpha_U, \nu, \lambda_U)}$ is the value of the noncentral t distribution at the α_U quantile with noncentrality parameter λ_U , respectively.

The confidence interval transformation principle becomes useful, because a confidence interval for the population noncentrality parameter is not generally of interest in and of itself. However, since there is a one-to-one monotonic relation between λ and κ , a confidence interval for λ can be transformed into a confidence interval for κ . If one rearranges Equation 5, it can be seen that

$$\kappa = \frac{\sqrt{N}}{\lambda}. \tag{7}$$

Therefore, transforming the limits of the confidence interval for λ , by dividing the limits by \sqrt{N} and then taking the inverse, confidence limits for κ can be obtained:

$$p \left[\left(\frac{\lambda_U}{\sqrt{N}} \right)^{-1} \leq \kappa \leq \left(\frac{\lambda_L}{\sqrt{N}} \right)^{-1} \right] = 1 - \alpha, \tag{8}$$

where p represents probability. Notice that the lower confidence limit for κ is obtained by dividing the upper confidence limit for λ by \sqrt{N} and that the upper confidence limit for κ is obtained by dividing the lower confidence limit for λ by \sqrt{N} . The reversal of limits is a necessary part of the transformation procedure in the present context.

As an example of forming a confidence interval for κ , an example is given for a situation that deals with Alzheimer’s disease patients. Volkow et al. (2002) studied the reduction of glucose metabolism in the brains of Alzheimer’s disease patients (as part of a larger study). The coefficient of variation across the entire cortex for the glucose metabolism of segmented cortical regions in the brains of

the 35 Alzheimer's patients was .271. Using Equation 6, the estimated noncentrality parameter ($\hat{\lambda}$) in this situation is 21.831 (i.e., $\sqrt{(35)/.271}$). A 95% confidence interval for the noncentrality parameter is given as

$$CI_{.95} = [16.282 \leq \lambda \leq 27.331],$$

where $CI_{.95}$ represents a 95% confidence interval. Converting the confidence limits for λ to κ by way of Equation 8 leads to the following confidence interval for κ :

$$CI_{.95} = [.216 \leq \kappa \leq .363].$$

Note that the confidence interval is *not* symmetric about k , where the lower confidence interval width is .055 (.271 - .216) and the upper confidence interval width is .092 (.363 - .271). In general, noncentral distributions are not symmetric, and confidence intervals based on noncentral distributions, even when $\alpha_L = \alpha_U$, tend to have different lower and upper confidence interval widths. Thus, confidence intervals for κ , which are based on the transformed confidence limits for the noncentral parameter, tend not to be symmetric about k .

AIPE for the Coefficient of Variation

When planning sample size, in order for the expected width of the obtained confidence interval to be sufficiently narrow for the population coefficient of variation, it is necessary to use an iterative process.⁷ Because the confidence interval width for κ is not symmetric (due to the nature of the noncentral t distribution), the desired width can pertain to the full confidence interval width, the lower width, or the upper width. Let $\kappa_L(\mathbf{X})$ be defined as the random lower confidence limit for κ and $\kappa_U(\mathbf{X})$ be defined as the random upper confidence limit for κ at the specified confidence level, where \mathbf{X} represents the observed data matrix on which the confidence interval is based. For notational ease, the lower and upper random confidence interval limits will be written as κ_L and κ_U , respectively, with the understanding that they are random because they are based on the observed data. The full width of the obtained confidence interval is thus given as

$$w = \kappa_U - \kappa_L, \quad (9)$$

the lower width of the obtained confidence interval is given as

$$w_L = k - \kappa_L, \quad (10)$$

and the upper width of the obtained confidence interval is given as

$$w_U = \kappa_U - k. \quad (11)$$

The goals of the research study will dictate for which confidence interval width sample size should be determined. In general, w will be the width of interest. Although the methods discussed are directly applicable to determining sample size for the lower or the upper confidence interval width (i.e., w_L or w_U), the focus of the present work is on the full confidence interval width (i.e., w). Let ω be defined as the desired confidence interval width, which is specified a priori by the researcher, much as the de-

sired degree of statistical power is chosen a priori when one is planning necessary sample size in a power analytic context (e.g., Cohen, 1988; Kraemer & Thieman, 1987; Lipsey, 1990; Murphy & Myers, 1998).

As is true with essentially all sample size planning procedures (e.g., power analysis), in order to plan sample size under the goals of AIPE, κ must be known or its value estimated. This complication is the same as almost all sample size planning procedures (e.g., power analysis). Conceptually, the way in which sample size is determined so that the expected width of the confidence interval for κ is sufficiently narrow is to substitute κ for k and systematically evaluate different sample sizes until the expected width of the confidence interval is no wider than desired (i.e., $E[w] \leq \omega$). Actually, since k systematically underestimates κ (recall Equation 3), basing the sample size planning procedure on κ instead of on $E[k]$ would lead to sample size estimates larger than necessary. For this reason, $\tilde{k}_{(\kappa, N)}$ from Equation 3 is substituted for κ when determining the necessary sample size so that the expected width is no larger than ω . Note that since $\tilde{k}_{(\kappa, N)}$ is in part a function of N , $\tilde{k}_{(\kappa, N)}$ will be updated for each iteration of the sample size procedure, as each iteration is based on a different sample size. This method of planning sample size is consistent with other sample size planning methods so that the expected width is sufficiently narrow (e.g., Guenther, 1981; Hahn & Meeker, 1991; Kelley & Maxwell, 2003; Kelley et al., 2003; Kelley & Rausch, 2006; Kupper & Hafner, 1989).

Although the overarching approach to sample size planning from an AIPE approach is not new (e.g., Mace, 1964), it has not been discussed much in the behavioral, educational, and social sciences. Furthermore, it has only recently been discussed for effect sizes that follow noncentral distributions (e.g., Kelley, 2007b; Kelley & Maxwell, in press; Kelley & Rausch, 2006; see also Algina & Olejnik, 2000, for a similar goal in a related context). Given the importance placed on effect sizes and confidence intervals in the literature, sample size planning from an AIPE approach will almost certainly increase in importance and frequency of usage.

Operationally, an algorithm that guarantees finding the appropriate sample size is to start at some minimal sample size, say N_0 , and determine $E[k | (\kappa, N_0)]$ so that $E[w | \tilde{k}_{(\kappa, N_0)}]$ can be determined. If the width is greater than desired, the sample size should be increased by one, and the confidence interval width determined again. This iterative process of increasing sample size and recalculating the expected confidence interval width should continue until $E[w | \tilde{k}_{(\kappa, N_i)}]$ is equal to or less than ω , where i represents the particular iteration number.

One issue that might not be obvious is that the method for planning sample size just discussed plans the necessary sample size so that the expected width of the confidence interval is sufficiently narrow; however, it does not guarantee that for any particular confidence interval the observed width will be sufficiently narrow. The confidence interval width, w , is a random variable that will fluctuate from sample to sample. The fact that sample size is determined so that $E[w]$ is no larger than κ implies that

roughly 50% of the sampling distribution of w s will be less than ω .⁸ A modified sample size procedure, discussed in the next section, can be implemented so that one can have a desired degree of assurance that the obtained w will be no larger than ω ; that is, a probabilistic statement can be incorporated into the sample size planning procedure that guarantees the obtainment of a sufficiently narrow confidence interval with some desired degree of assurance.

Ensuring a Confidence Interval No Wider Than Desired With a Specified Degree of Assurance

One property of the method for forming confidence intervals using the noncentral t distribution for κ is that as k increases, so does the width of the confidence interval for κ . This implies that when k is larger than the κ on which the standard sample size procedure is based, w will be larger than ω . In order to avoid obtaining a k larger than the value the sample size procedure is based on with some specified degree of assurance, denoted γ , and thus a w wider than ω , a modified sample size procedure can be used that substitutes for κ from the standard procedure κ_γ , where k will not exceed κ_γ 100% of the time. That is, κ_γ is the value at the γ th quantile from the particular sampling distribution of k . Thus, given the specified degree of assurance, κ_γ is the largest plausible value of k expected to be obtained with probability γ for a particular sample size and κ value. The value $1 - \gamma$ thus represents the probability that k will be greater than κ , which implies that w will be greater than ω for the particular sample size.

When κ_γ is found based on the necessary sample size from the standard procedure, κ_γ can be substituted for κ in the standard procedure and sample size determined in the same manner as before. κ_γ is obtained by transforming the γ th quantile from a noncentral t distribution with noncentrality parameter λ and ν degrees of freedom. The value λ_γ is the value of the particular noncentral t distribution that satisfies

$$\int_{-\infty}^{\lambda_\gamma} f(t_{(\nu;\lambda)}) dt = \gamma, \tag{12}$$

where $f(t_{(\nu;\lambda)})$ represents the noncentral t distribution probability function, and ν is based on the sample size from the standard procedure.

The rationale for replacing κ from the standard sample size procedure with κ_γ is so that w will exceed ω no more than $(1 - \gamma)$ 100% of the time. Since

$$p(k \leq \kappa_\gamma) \geq \gamma, \tag{13}$$

no more than $(1 - \gamma)$ 100% percent of the time will w be greater than ω :

$$p(w \geq \omega) \leq 1 - \gamma. \tag{14}$$

Thus, the modified sample size procedure ensures that the observed confidence interval will not be wider than ω with probability no less than γ .⁹ The method of determining the modified sample size is consistent in theory with other methods of sample size planning in order to attach a probabilistic statement to the confidence interval width being sufficiently narrow (e.g., Guenther, 1981; Hahn &

Meeker, 1991; Kelley & Maxwell, 2003; Kelley et al., 2003; Kelley & Rausch, 2006; Kupper & Hafner, 1989).

Tables of Necessary Sample Size

Although the Appendix provides information on implementing the methods proposed and discussed in order to obtain sufficiently narrow confidence intervals for any combination of κ , ω , α , and γ using the Methods for the Behavioral, Educational, and Social Sciences (MBESS; Kelley, 2007b, 2007c) R package (R Development Core Team, 2007), tables of selected conditions are provided. The tables are not meant to include all potentially interesting conditions; rather, they are intended to provide researchers (1) a convenient way to plan sample size when the situation of interest is approximately that included in the table, and (2) a way to illustrate the relation between κ , ω , γ , α , and necessary sample size. The tabled values are based on κ values thought to be useful for areas of the behavioral, educational, and social sciences.

Sample size is tabled for κ values of 0.05 to 0.50 by 0.05, ω values of .01, .025 to .20 by .025, a subtable where the expected value of w equals ω , along with subtables with γ values of .80 and .99, each for confidence interval coverages of .90 (Table 1), .95 (Table 2), and .99 (Table 3). Each of the conditions is crossed with all other conditions in a factorial manner, and thus there are a total of 810 situations ($10 \times 9 \times 3 \times 3$) for planning an appropriate sample size.¹⁰

Any time a necessarily positive quantity follows a normal distribution, it is unlikely that k values larger than $\frac{1}{2}$ can be obtained. Values of k larger than $\frac{1}{2}$ imply that the standard deviation is more than three times larger than the mean, which would further imply that the lower end of the distribution would be expected to contain some proportion less than zero. For example, if $\mu = 30$ and $\sigma = 10$ from a normal distribution, and thus $\kappa = \frac{1}{2}$, .135% of the distribution would fall below zero. This small proportion of expected negative scores, when the normally distributed quantity is necessarily positive, may be ignorable in this situation; but as κ increases arbitrarily large, the proportion of expected negative scores increases arbitrarily close to 50% of the distribution. The same discussion applies to quantities that are necessarily negative and that follow a normal distribution. Thus, for necessarily positive quantities k values tend to be less than $\frac{1}{2}$. Values of k greater than $\frac{1}{2}$ tend to arise when the distribution consists of both negative and positive values, or when the distribution consists only of positive or negative values but is not normally distributed.

Suppose a researcher is interested in estimating a coefficient of variation that has a corresponding confidence for κ that is sufficiently narrow. After a literature review of studies that examined a similar phenomenon under similar conditions, it was hypothesized that the population coefficient of variation was .25. With the desire to obtain a 99% confidence interval that excludes 0.20 and 0.30, the researcher sets ω to 0.10.¹¹ Application of the methods leads to a necessary sample size of $N = 99$. This sample size is contained in the first subtable of Table 3, specifically (from the top left) five cells down (the $\omega = 0.10$ row) and five cells over (the $\kappa = 0.25$ column).

Table 1
Necessary Sample Size for 90% Confidence Intervals
for the Coefficient of Variation in Selected Situations

ω	κ									
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$E[w]$										
.010	141	557	1,277	2,342	3,810	5,752	8,258	11,434	15,401	20,298
.025	26	93	208	379	614	925	1,326	1,834	2,469	3,253
.050	9	27	56	98	157	235	335	463	622	818
.075	6	14	26	45	71	106	151	207	278	365
.100	5	9	17	27	42	62	87	119	159	209
.125	—	7	12	19	29	41	57	78	104	136
.150	—	6	10	14	21	30	41	56	74	96
.175	—	5	8	12	17	23	32	42	56	72
.200	—	5	7	10	14	19	25	33	44	56
$\gamma = .80$										
.010	154	585	1,322	2,406	3,896	5,865	8,402	11,616	15,627	20,576
.025	31	104	226	404	648	970	1,384	1,907	2,560	3,365
.050	11	32	64	111	174	257	364	499	668	874
.075	7	16	31	53	82	121	170	232	309	404
.100	6	11	20	33	50	73	102	138	182	237
.125	—	9	15	23	35	50	69	127	122	159
.150	—	7	12	18	26	37	51	68	89	115
.175	—	6	10	14	21	29	40	53	69	89
.200	—	6	8	12	17	24	32	43	55	71
$\gamma = .99$										
.010	180	638	1,405	2,523	4,053	6,070	8,664	11,945	16,036	21,077
.025	42	126	261	453	714	1,055	1,493	2,044	2,730	3,573
.050	16	43	82	137	209	303	423	573	758	985
.075	10	25	45	72	108	154	212	285	374	483
.100	8	18	31	49	71	100	136	180	235	301
.125	—	14	23	36	52	73	98	129	167	213
.150	—	11	19	29	41	57	77	100	129	164
.175	—	10	16	24	34	47	63	82	105	133
.200	—	9	14	21	29	40	54	70	89	113

Note— κ is the population coefficient of variation, γ is the desired degree of assurance of achieving a confidence interval for κ no wider than desired, ω is the desired full confidence interval width, and $E[w]$ is the expected confidence interval width (i.e., when γ is not specified).

Realizing that an N of 99 will lead a sufficiently narrow confidence interval only about half of the time, the researcher incorporates an assurance (i.e., γ) of .99. A γ of .99 implies that the width of the 99% confidence interval will be greater than desired (i.e., 0.10) no more than 1% of the time. From the third subtable of Table 3 (again five cells down and five cells over from the top left), it can be seen that the modified sample size procedure yields a necessary sample size of 141. Using a sample size of 141 will thus provide 99% assurance that the obtained confidence interval for κ will be no wider than 0.10 units.

A summary of the results contained in the table is provided. As can be seen, holding all other factors constant, (1) larger values of κ , (2) smaller values of ω , (3) larger values of γ , and (4) larger values of the confidence interval coverage (i.e., $1 - \alpha$) all lead to larger necessary sample sizes. These findings are intuitively reasonable, in that a quantity in a population with more variability is estimated as κ increases, implying larger standard deviations relative to the mean (or smaller means relative to the standard deviation). Smaller values of ω imply that a larger sample size is necessary in order to reduce the standard error and increase the degrees of freedom, so that the confidence interval becomes narrower. Having more assurance that the confidence inter-

val will be sufficiently narrow (e.g., 99% certain compared with 80% certain) also requires a larger sample size, because increasing the probabilistic component implies that it will be more difficult to achieve the goal satisfactorily. Similarly, as the confidence interval coverage increases (i.e., a decrease in α), for a desired confidence interval width the sample size also increases. The reverse is also true: (1) a decrease in κ (e.g., by reducing variability in the sample), (2) an increase in ω , (3) smaller values of γ , and (4) smaller confidence interval coverages all lead to smaller sample sizes.

Sensitivity Analyses

In almost all situations, κ will be unknown; yet κ must be specified in order to plan the appropriate sample size. This is the conundrum for most sample size planning procedures. One question that arises is “What are the effects on the width of the confidence interval if the value of κ specified is *not* equal to the population value?” This is an analogous problem when planning sample size in a power analytic context, where sensitivity analyses are often suggested. Such sensitivity analyses are also recommended in the context of AIPE.

The idea of a sensitivity analysis in the present context is to assess the effects of misspecifying the population

Table 2
Necessary Sample Size for 95% Confidence Intervals
for the Coefficient of Variation in Selected Situations

ω	κ									
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$E[w]$										
.010	199	790	1,812	3,325	5,408	8,166	11,724	16,233	21,866	28,819
.025	37	131	295	537	871	1,312	1,882	2,603	3,505	4,618
.050	13	37	78	139	222	333	475	656	882	1,160
.075	8	18	36	63	100	149	213	296	396	520
.100	6	12	22	38	59	87	123	168	225	294
.125	5	9	16	26	40	58	81	110	146	191
.150	5	8	13	20	29	42	58	78	104	135
.175	—	7	10	16	23	32	44	59	78	101
.200	—	6	9	13	18	26	35	47	61	79
$\gamma = .80$										
.010	215	823	1,866	3,401	5,511	8,300	11,896	16,450	22,136	29,150
.025	42	144	316	567	911	1,366	1,950	2,690	3,613	4,751
.050	16	43	88	153	242	360	510	700	936	1,227
.075	9	22	43	73	113	167	238	325	433	565
.100	7	14	27	45	69	100	140	190	253	328
.125	6	11	19	31	47	68	94	127	169	219
.150	5	9	15	24	35	50	69	93	122	158
.175	—	8	12	19	28	39	54	72	94	121
.200	—	7	11	16	23	32	43	57	75	97
$\gamma = .99$										
.010	246	886	1,964	3,540	5,698	8,544	12,207	16,841	22,621	29,745
.025	55	170	357	625	989	1,466	2,079	2,852	3,814	4,996
.050	22	57	110	184	283	413	578	785	1,042	1,357
.075	13	31	58	94	142	205	286	385	507	656
.100	9	22	39	62	92	131	179	239	311	401
.125	8	17	30	46	67	93	127	168	219	281
.150	7	14	23	36	52	72	98	129	167	212
.175	—	12	20	30	43	59	79	104	134	170
.200	—	10	17	26	37	50	67	87	112	142

Note— κ is the population coefficient of variation, γ is the desired degree of assurance of achieving a confidence interval for κ no wider than desired, ω is the desired full confidence interval width, and $E[w]$ is the expected confidence interval width (i.e., when γ is not specified).

value of κ directly (or indirectly by misspecifying μ and/or σ) on the typical confidence interval width and the proportion of sufficiently narrow confidence intervals. For example, suppose κ is specified to be 0.20 for purposes of the sample size planning procedure, yet in actuality the value of κ is 0.30. The question that a sensitivity analysis addresses in this situation concerns the properties of the confidence interval width for k when using the sample size necessary when κ is based on 0.20, when in fact the true value of κ is 0.30.

The way in which a sensitivity analysis can be implemented is by performing a Monte Carlo simulation study in a population where the value of κ is set to be the true value but the sample size used is based on the misspecified κ value. Data is generated when all assumptions hold (i.e., normality and independence of observations) and a confidence interval is calculated for each of the large number of generated data sets (e.g., 10,000). As the simulation study progresses, information of interest is recorded for subsequent analysis (e.g., mean confidence interval width, proportion of confidence intervals less than the desired width, etc.). In some cases, the effect of misspecifying κ a minimal amount is trivial, whereas other times it can be quite large. The combination of all factors needs

to be known before a statement about the effects of the misspecification of κ can be known. In general, properties of misspecification are not known analytically; hence the need for a Monte Carlo simulation study. The Appendix provides information on how a sensitivity analysis can be implemented in the AIPE context for the coefficient of variation using MBESS.

Discussion

If a point estimate is of interest, the confidence limits that bracket the population quantity also should be. Regardless of the value of a point estimate, the population quantity will almost certainly differ. Since it is the population value, not a sample estimate, that is ultimately of interest, the limits of the confidence interval are arguably more important than is the point estimate itself. The limits of the confidence interval bracket what can be considered plausible values of the population parameter with some specified level of confidence. As the width of the interval narrows, holding constant the level of confidence, more and more values are excluded and thus are no longer considered plausible. The width of the confidence interval is therefore a way through which the accuracy of the parameter estimate can be operationalized (e.g., Neyman,

Table 3
Necessary Sample Size for 99% Confidence Intervals
for the Coefficient of Variation in Selected Situations

ω	κ									
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
	E[w]									
.010	342	1,362	3,129	5,742	9,340	14,102	20,248	28,037	37,766	49,774
.025	62	225	508	926	1,502	2,265	3,248	4,495	6,052	7,974
.050	21	63	134	238	383	573	820	1,132	1,522	2,003
.075	11	30	64	111	175	260	370	509	683	897
.100	9	19	37	63	99	147	209	287	389	510
.125	7	14	26	43	66	97	136	187	250	327
.150	6	12	20	32	49	70	98	134	177	232
.175	6	10	16	25	38	54	74	100	133	173
.200	5	9	14	21	30	43	59	79	104	135
	$\gamma = .80$									
.010	363	1,407	3,199	5,842	9,475	14,279	28,321	50,209	38,120	82,817
.025	69	242	536	966	1,556	2,335	3,339	4,609	6,194	8,149
.050	25	71	147	258	409	609	865	1,189	1,593	2,090
.075	13	35	73	123	193	283	400	547	730	956
.100	10	23	43	72	112	164	231	320	425	554
.125	8	17	30	50	76	110	154	210	279	363
.150	7	14	23	38	57	82	113	153	202	262
.175	6	11	19	30	44	63	87	117	154	199
.200	6	10	16	25	36	51	70	93	122	157
	$\gamma = .99$									
.010	404	1,489	3,327	6,023	9,719	14,597	20,881	28,832	38,753	50,986
.025	86	276	589	1,041	1,656	2,466	3,506	4,819	6,454	8,467
.050	32	88	174	297	461	677	952	1,298	1,728	2,256
.075	19	49	91	150	229	331	460	623	824	1,071
.100	13	31	58	93	141	201	279	379	498	644
.125	11	24	43	68	100	142	195	260	341	439
.150	9	19	34	53	77	109	148	196	255	327
.175	8	17	28	44	63	87	118	156	202	257
.200	8	15	24	37	53	73	98	128	166	211

Note— κ is the population coefficient of variation, γ is the desired degree of assurance of achieving a confidence interval for κ no wider than desired, ω is the desired full confidence interval width, and E[w] is the expected confidence interval width (i.e., when γ is not specified).

1937). Since the value of interest is the population value, designing a study to obtain an accurate estimate should be a top concern for researchers.

At present, it is not clear to what extent violations of the assumption of normality and independent observations will have on the confidence interval coverage and/or the necessary sample size from the procedure developed. Because the sample size planning procedure is based on the parametric confidence interval procedure, in situations in which the confidence interval procedure fails to provide appropriate coverage (i.e., when the assumptions of the procedure are violated), the sample size obtained from the procedure will probably not be optimal. Thus, in order for the sample size from the procedure to be appropriate, the assumptions need to be satisfied. At present, the robustness of the confidence interval procedure and the sample size planning method are unknown. Issues of robustness, and appropriateness of sample size in violations of assumptions, are certainly areas that could benefit from additional research.

Generally, when the term “sample size planning” is used, it is taken to mean power analysis, where the goal of the procedure is to be able to reject the null hypothesis with some specified probability. Note that the power

analytic and the AIPE approach to sample size planning have fundamentally different goals. Necessary sample size to achieve a statistically significant estimate (i.e., power analysis) may be much different than necessary sample size to achieve a narrow confidence interval (i.e., AIPE). Depending on the particular situation, the power analytic approach or the AIPE approach could require a larger sample size (see Kelley & Maxwell, 2003; Kelley et al., 2003; Kelley & Rausch, 2006, for comparisons of necessary sample sizes for the power analytic and AIPE approach in difference situations).

From the outset, this article has made the assumption that the coefficient of variation can be an important and helpful quantity when trying to understand relative variability. If the coefficient of variation is indeed of interest, the confidence interval for the population coefficient of variation also should be. Since from a probabilistic perspective a wide confidence interval illustrates an estimate with a low degree of expected accuracy—an obviously undesirable situation—an effort should be made to obtain an estimate with a high degree of expected accuracy whenever possible. When planning studies and selecting an appropriate sample size in situations where the coefficient of variation is of interest, it is hoped that researchers take

into consideration this article's discussion of confidence interval width and the appropriate sample size.

AUTHOR NOTE

This work was sponsored in part by a Proffitt Fellowship for Educational Research. Correspondence concerning this article should be addressed to K. Kelley, Inquiry Methodology Program, Indiana University, 201 North Rose Avenue, Bloomington, IN 47405 (e-mail: kkkiii@indiana.edu).

REFERENCES

- ALGINA, J., & OLEJNIK, S. (2000). Determining sample size for accurate estimation of the squared multiple correlation coefficient. *Multivariate Behavioral Research*, *35*, 119-136.
- BABKOFF, H., KELLY, T. L., & NAITOH, P. (2001). Trial-to-trial variance in choice reaction time as a measure of the effect of stimulants during sleep deprivation. *Military Psychology*, *13*, 1-16.
- BEDEIAN, A. G., & MOSSHOLDER, K. W. (2000). On the use of the coefficient of variation as a measure of diversity. *Organizational Research Methods*, *3*, 285-297.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- COHEN, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997-1003.
- DINGES, D. F., & KRIBBS, N. B. (1991). Performance while sleepy: Effects of experimentally-induced sleepiness. In T. H. Monk (Ed.), *Sleep, sleepiness, and performance* (pp. 97-128). New York: Wiley.
- EFRON, B., & TIBSHIRANI, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall/CRC.
- FRITH, U., & FRITH, C. (2001). The biological basis of social interaction. *Current Directions in Psychological Science*, *10*, 151-155.
- GUENTHER, W. C. (1981). Sample size formulas for normal theory T tests. *American Statistician*, *35*, 243-244.
- HAHN, G., & MEEKER, W. (1991). *Statistical intervals: A guide for practitioners*. New York: Wiley.
- HALDANE, J. B. S. (1955). The measurement of variation. *Evolution*, *9*, 484.
- HAYASHI, R. (2000). Correlation between coefficient of variation of choice reaction time and components of event-related potentials (P300): Effect of benzodiazepine. *Journal of the Neurological Sciences*, *178*, 52-56.
- HUNTER, J. E., & SCHMIDT, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- JOHNSON, N. L., KOTZ, S., & BALAKRISHNAN, N. (1995). *Continuous univariate distributions* (2nd ed., Vol. 2). New York: Wiley.
- JOHNSON, N. L., & WELCH, B. L. (1940). Applications of the noncentral t distribution. *Biometrika*, *31*, 362-389.
- KELLEY, K. (2007a). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software*, *20*, 1-24.
- KELLEY, K. (2007b). Methods for the Behavioral, Educational, and Social Sciences (MBESS) [Computer software and manual]. Retrieved from www.cran.r-project.org/.
- KELLEY, K. (2007c). Methods for the behavioral, educational, and social sciences: An R package. *Behavior Research Methods*, *39*, 979-984.
- KELLEY, K. (2007d). *Sample size planning for the squared multiple correlation coefficient: Accuracy in parameter estimation via narrow confidence intervals*. Manuscript submitted for publication.
- KELLEY, K., & MAXWELL, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, *8*, 305-321.
- KELLEY, K., & MAXWELL, S. E. (in press). Sample size planning for multiple regression: Power and accuracy for omnibus and targeted effects. In J. Brannon, P. Alasuutari, & L. Bickman (Eds.), *Sage handbook of social research methods*. Thousand Oaks, CA: Sage.
- KELLEY, K., MAXWELL, S. E., & RAUSCH, J. R. (2003). Obtaining power or obtaining precision: Delineating methods of sample size planning. *Evaluation & the Health Professions*, *26*, 258-287.
- KELLEY, K., & RAUSCH, J. R. (2006). Sample size planning for the standardized mean difference: Accuracy in parameter estimation via narrow confidence intervals. *Psychological Methods*, *11*, 363-385.
- KIRK, R. (2001). Promoting good statistical practice: Some suggestions. *Educational & Psychological Measurement*, *61*, 213-218.
- KOSSLYN, S. M., CACIOPPO, J. T., DAVIDSON, R. J., HUGDAHL, K., LOVALLO, W. R., SPIEGEL, D., & ROSE, R. (2002). Bridging psychology and biology: The analysis of individuals in groups. *American Psychologist*, *57*, 341-351.
- KRAEMER, H. C., & THIEMANN, S. (1987). *How many subjects?: Statistical power analysis in research*. Newbury Park, CA: Sage.
- KUPPER, L. L., & HAFNER, K. B. (1989). How appropriate are popular sample size formulas? *The American Statistician*, *43*, 101-105.
- LIPSEY, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.
- MACE, A. E. (1964). *Sample size determination*. New York: Reinhold.
- MCKAY, A. T. (1932). Distribution of the coefficient of variation and the extended "t" distribution. *Journal of the Royal Statistical Society*, *95*, 695-698.
- MEEHL, P. E. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 393-426). Mahwah, NJ: Erlbaum.
- MONCHAR, P. H. (1981). Regional educational inequality and political instability. *Comparative Education Review*, *25*, 1-12.
- MURPHY, K. R., & MYORS, B. (1998). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Mahwah, NJ: Erlbaum.
- NEYMAN, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society A*, *236*, 333-380.
- R DEVELOPMENT CORE TEAM (2007). R: A language and environment for statistical computing [Computer software and manual], R Foundation for Statistical Computing. Retrieved from www.r-project.org.
- REED, G. F., LYNN, F., & MEADE, B. D. (2002). Use of coefficient of variation in assessing variability of quantitative assays. *Clinical & Diagnostic Laboratory Immunology*, *9*, 1235-1239.
- ROZEBOOM, W. W. (1966). *Foundations of the theory of prediction*. Homewood, IL: Dorsey.
- SALMON, P., & HALL, G. M. (1997). A theory of postoperative fatigue: An interaction of biological, psychological, and social processes. *Pharmacology Biochemistry & Behavior*, *56*, 623-628.
- SCHMIDT, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, *1*, 115-129.
- SHAFIR, S. (2000). Risk-sensitivity foraging: The effect of relative variability. *Oikos*, *88*, 663-669.
- SHERET, M. (1984). Note on methodology: The coefficient of variation. *Comparative Education Review*, *28*, 467-476.
- SMITHSON, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational & Psychological Measurement*, *61*, 605-632.
- SOKAL, R. R., & BRAUMANN, C. A. (1980). Significance tests for coefficients of variation and variability profiles. *Systematic Zoology*, *29*, 50-66.
- STEIGER, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, *9*, 164-182.
- STEIGER, J. H., & FOULADI, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical methods. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221-257). Mahwah, NJ: Erlbaum.
- TASK FORCE ON REPORTING OF RESEARCH METHODS IN AERA PUBLICATIONS (2006). *Standards for reporting on empirical social science research in AERA publications*. Washington, DC: American Educational Research Association.
- THOMPSON, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, *31*, 25-32.
- VELLEMAN, P. F., & WILKINSON, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *American Statistician*, *47*, 65-72.
- VOLKOW, N. D., ZHU, W., FELDER, C. A., MUELLER, K., WELSH, T. F.,

- WANG, G.-J., & DE LEON, M. J. (2002). Changes in brain functional homogeneity in subjects with Alzheimer's disease. *Psychiatry Research: Neuroimaging*, **114**, 39-50.
- WEBER, E. U., SHAFIR, S., & BLAIS, A.-R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review*, **111**, 430-445.
- WILKINSON, L., & THE AMERICAN PSYCHOLOGICAL ASSOCIATION TASK FORCE ON STATISTICAL INFERENCE (1999). Statistical methods in psychology: Guidelines and explanations. *American Psychologist*, **54**, 594-604.
- WILLIAMS, K. Y., & O'REILLY, C. A., III (1998). Demography and diversity in organizations: A review of 40 years of research. *Research in Organizational Behavior*, **20**, 77-140.

NOTES

1. In fact, "many parts of the literature" can be identified with specific journals, where the overarching goal of the journals combines various aspects of biological systems and processes with psychological systems and processes (e.g., *Behavioral Ecology*; *Behavioural Brain Research*; *Behavioural Pharmacology*; *Biological Psychology*; *Brain Research*; *Developmental Psychobiology*; *European Neuropsychopharmacology*; *Genes, Brain & Behavior*; *Hormones & Behavior*; *International Journal of Psychophysiology*; *Journal of Behavioral Medicine*; *Journal of Clinical & Experimental Neuropsychology*; *Neurobiology of Learning & Memory*; *Physiology & Behavior*; *Pharmacology Biochemistry & Behavior*; and *Psychoneuroendocrinology*, to list a nonexhaustive set of peer reviewed journals).

2. Using the work of Monchar (1981) as partial motivation, and improving the quantitative measure of inequities as another, Sheret (1984) discussed using a weighted coefficient of variation when the number of observations within groups differed across a set of groups for which the coefficient of variation was to be calculated.

3. It should be noted that there is some debate as to whether or not the coefficient of variation should be computed only for nonnegative ratio scaled data (e.g., Bedeian & Mossholder, 2000). Although there is a considerable literature on the subject of what should and should not be done to a set of numbers (i.e., the measurement scale debate; e.g., Velleman & Wilkinson, 1993, for a review), the coefficient of variation can certainly be computed whenever the mean and standard deviation are available. Of course, whether or not the coefficient of variation is a meaningful quantity depends on the particular situation and the question of interest. Regardless of which side of the measurement scales debate one comes down on, the methods discussed in the article can be used whenever there is an interest in the coefficient of variation.

4. Assuming that the assumptions of the model are met, the correct model is fit, and observations are randomly sampled, $(1 - \alpha)$ is the probability that any given confidence interval from a collection of confidence intervals calculated under the same circumstances will contain the population parameter of interest. However, it is not true that a specific confidence interval is correct with $(1 - \alpha)$ probability, since a computed confidence interval either does or does not contain the value of the parameter. The confidence interval procedure refers to the infinite number of confidence intervals that could theoretically be constructed, and the $(1 - \alpha)100\%$ of those confidence intervals that correctly bracket the population parameter of interest (see Hahn & Meeker, 1991, for a technical review of confidence interval formation and interpretation). Although the meaning of confidence intervals has been given from a frequentist perspective, the methods discussed in the article are equally applicable under the Bayesian perspective of confidence interval interpretation.

5. Methods for the Behavioral, Educational, and Social Sciences (MBESS; Kelley, 2007a; Kelley, in press) is an R package (R Develop-

ment Core Team, 2007) that implements all of the methods discussed in the article, and many more. Both R and MBESS are Open Source and freely available. R and MBESS can be obtained from the Internet at the Comprehensive R Archival Network: cran.r-project.org. The specific MBESS Web page on the Comprehensive R Archival Network is cran.r-project.org/src/contrib/Descriptions/MBESS.html, where the most current version and the documentation can be found (note that this Internet address is case sensitive). The Appendix provides details on MBESS and example code illustrating how each of the methods discussed can be easily implemented. Researchers interested in implementing the methods and techniques discussed in the article can thus readily do so using MBESS.

6. It should be pointed out that the exact analytic confidence interval based on the noncentral t distribution is not the only way to construct a legitimate confidence interval for κ . One possibility is to make use of the bootstrap technique, a nonparametric method that does not make the assumption of normality (e.g., Efron & Tibshirani, 1993). Although the bootstrap is a perfectly legitimate alternative to the analytic approach, the sample size planning methods have been developed based on the analytic approach. Of course, one could plan sample size based on the analytic approach and compute confidence intervals with the bootstrap approach. However, if normality held in the population, bootstrap confidence intervals would tend to be wider than the normal theory analytic approach confidence intervals, on which the sample size planning procedure was based (Efron & Tibshirani, 1993). Thus, using the methods developed here for sample size planning based on the population being normal, then using the bootstrap approach to confidence interval formation in situations where normality is likely violated, will probably not yield optimal sample size planning values.

7. Purely analytic solutions do not generally exist for finding quantiles from the noncentral t distribution, and thus noncentral routines that deal with noncentral t distributions are essentially always iterative in nature. Given modern software, this poses no problem. However, the lack of computing power once motivated a literature of approximate ways to obtain necessary values from a noncentral t distribution. The difficulty in finding the necessary confidence limits led to many tri-entry (probability, degrees of freedom, and noncentrality parameter) tables (see a review of such tables in Johnson, Kotz, & Balakrishnan, 1995, chap. 31) that will almost certainly yield approximate results in any applied situation.

8. Typically just more than 50% of the sampling distribution of w will be less than ω , because $E[w|\hat{k}_{(\kappa, N)}]$ will typically be less than ω . Although ideally $E[w|\hat{k}_{(\kappa, N)}] = \omega$, because sample size changes in whole numbers, in order for $E[w|\hat{k}_{(\kappa, N)}]$ to be no larger than ω , $E[w|\hat{k}_{(\kappa, N)}]$ will typically be just less than ω . Coupling this fact with the positively skewed nature of the noncentral t distribution (when the noncentral parameter is positive) typically leads to just more than 50% of the sampling distribution of w being less than ω .

9. Notice that Equations 13 and 14 are inequalities rather than equalities (see also Note 8). The reason for the lack of perfect equality is that the confidence interval width changes as a function of N , which follows a step function. This is the case because in practice sample size consists of only whole numbers; thus, the change in N is not a continuous and smooth function. In order to ensure that the confidence interval is no wider than ω , $p(w \geq \omega)$ will almost always be less than γ . Although Equations 13 and 14 theoretically should be equalities, maintaining sample size at only whole numbers implies a need for inequalities.

10. Note that 20 cells out of the 810 cells could not be computed because of the large value of ω relative to the small value of κ .

11. Although this hypothetical researcher's belief seems reasonable regarding the hope of excluding 0.20 and 0.30, due to the positively skewed nature of the noncentral t distribution (when $\lambda > 0$), an expected confidence interval width of 0.10 would more likely exclude 0.20 than it would 0.30.

APPENDIX

Using MBESS to Implement the Methods

All of the methods and procedures discussed and the algorithms presented can easily be implemented in the Methods for the Behavioral, Educational, and Social Sciences (Kelley, 2007a, 2007b, 2007c) R package (R Development Core Team, 2007). This appendix provides a brief overview of the way in which the necessary MBESS functions can be used when the coefficient of variation is of interest. Those not familiar with R will see that R is a command driven language. The commands (which are case sensitive) are input directly into the R console. Unlike some programming languages, R executes code line by line sequentially. R code that is directly executable in this appendix is preceded with `R>`, where `R>` is used to illustrate the R command prompt where commands are input and then executed (by entering the command). Both R and MBESS are Open Source and thus freely available. R and the MBESS package are available at the following Internet address for all commonly used operating systems: cran.r-project.org. The specific Internet address for MBESS is cran.r-project.org/src/contrib/Descriptions/MBESS.html. Because MBESS is an optional package, it must be loaded with each new R session where its routines will be used. Packages in R are loaded with the `library()` command, which is illustrated with MBESS as follows:

```
R> library(MBESS).
```

Confidence Intervals for a Noncentral t Parameter

For constructing confidence intervals for the noncentrality parameter from a noncentral t distribution, the `conf.limits.nct()` function can be used. The lower and upper critical values from the noncentral t distribution are returned by specifying the following arguments in the `conf.limits.nct()` function:

```
R> conf.limits.nct(ncp =  $\hat{\lambda}$ , df= $\nu$ , conf.level=  $1 - \alpha$ ),
```

where `ncp` is the estimated noncentrality parameter, `df` is the degrees of freedom for the particular situation, and `conf.level` is the desired level of confidence (i.e., $1 - \alpha$). For the Volkow et al. (2002) example discussed previously, the function would be implemented as

```
R> conf.limits.nct(ncp=21.831, df=34, conf.level=.95).
```

Confidence Intervals for the Coefficient of Variation

Given the one-to-one relation between λ and κ , and the confidence interval transformation principle previously discussed, the confidence limits for κ can be found by transforming the confidence limits for λ given the relation specified in Equation 7 and replacing what was the upper limit with what was the lower limit (and vice versa). Alternatively and simpler, the `ci.cv()` function can be used directly to determine the confidence limits for κ . The lower and upper critical value from Volkow et al. (2002) discussed previously are returned using the following specifications

```
R> ci.cv(cv=.271, n=35, conf.level=.95),
```

where `cv` is the observed coefficient of variation (i.e., k), `n` is the sample size, and `conf.level` is the desired level of confidence (i.e., $1 - \alpha$).

Planning Sample Size for the Coefficient of Variation

Using the example provided in the Tables of Necessary Sample Size section, where κ was set to .25 and ω set to .10 for a 99% confidence interval, the `ss.aipe.cv()` function can be used. The way in which the `ss.aipe.cv()` can be used so that the expected width is sufficiently narrow is given as

```
R> ss.aipe.cv(C.of.V=.25, width=.10, conf.level=.99),
```

where `C.of.V` is the population coefficient of variation (i.e., κ), `width` is the desired confidence interval width, and `conf.level` is the confidence level (i.e., $1 - \alpha$). Implementation of this function yields a necessary sample size of 99 (as reported in the text and in Table 3).

The way in which the desired degree of assurance can be used in the function is by specification of assurance, which is an optional argument in the `ss.aipe.cv()` function. The function would thus be specified as

```
R> ss.aipe.cv(C.of.V=.25, width=.10, conf.level=.99, assurance=.99),
```

which returns necessary sample size of 141 (as reported in the text and in Table 3).

Sensitivity Analysis for the Coefficient of Variation Given the Goals of AIPE

Sensitivity analysis to assess the effect of misspecifying κ on the width of the confidence interval can be performed with the `ss.aipe.cv.sensitivity()` function. The function `ss.aipe.cv.sensitivity()` allows one to specify the true population κ and an estimated but incorrect κ value, so that the effect of misspecifying κ on the width of the obtained confidence intervals can be empirically determined. The function performs a simulation where the empirical findings regarding the width of the confidence interval can be determined.

APPENDIX (Continued)

The results of the simulation within `ss.aipe.cv.sensitivity()` can be very helpful for determining how discrepant an incorrectly specified value of κ can be from κ itself in order to still have an acceptably narrow confidence interval for κ . The `ss.aipe.cv.sensitivity()` function can be specified as

```
R> ss.aipe.cv.sensitivity(True.C.of.V, Estimated.C.of.V, width,  
                          conf.level, G),
```

where `True.C.of.V` and `Estimated.C.of.V` are the true and the estimated κ values, `width` is the desired confidence interval width, `assurance` is the desired degree of assurance, `conf.level` is the desired confidence level (i.e., $1 - \alpha$), and `G` is the number of replications that take place within the simulation study (e.g., `G = 10,000`). Instead of specifying `Estimated.C.of.V`, a particular sample size can be specified using `Specified.N`, so that the properties of the confidence interval can be readily determined for a particular κ value at a specific sample size. A value of γ can also be specified with the addition of the optional `assurance` argument. The output of the function provides a thorough summary of the results, optionally with the full set of results from the Monte Carlo simulation study for the other analyses not implemented in the function can be performed on the results (e.g., visualization techniques).

(Manuscript received August 30, 2006;
revision accepted for publication February 1, 2007.)