

STATISTICAL POWER OF STUDIES EXAMINING THE COGNITIVE EFFECTS OF SUBTHALAMIC NUCLEUS DEEP BRAIN STIMULATION IN PARKINSON'S DISEASE

Steven Paul Woods¹, Julie D. Rippeth², Emily Conover¹, Catherine L. Carey³, Thomas D. Parsons⁴, and Alexander I. Tröster⁵

¹*Department of Psychiatry, University of California, San Diego, CA,* ²*Department of Psychiatry, Walter Reed Army Medical Center, Washington, D.C.,* ³*Department of Psychiatry and Behavioral Sciences, Duke University, NC; Joint Doctoral Program in Clinical Psychology, University of California, San Diego, and San Diego State University, CA,* ⁴*Department of Psychology, Fuller Theological Seminary, CA; Department of Neurology, University of North Carolina, Chapel Hill, NC, and* ⁵*Department of Neurology, University of North Carolina, Chapel Hill, NC*

It has been argued that neuropsychological studies generally possess adequate statistical power to detect large effect sizes. However, low statistical power is problematic in neuropsychological research involving clinical populations and novel interventions for which available sample sizes are often limited. One notable example of this problem is evident in the literature regarding the cognitive sequelae of deep brain stimulation (DBS) of the subthalamic nucleus (STN) in persons with Parkinson's disease (PD). In the current review, a post hoc estimate of the statistical power of 30 studies examining cognitive effects of STN DBS in PD revealed adequate power to detect substantial cognitive declines (i.e., very large effect sizes), but surprisingly low estimated power to detect cognitive changes associated with conventionally small, medium, and large effect sizes. Such wide spread Type II error risk in the STN DBS cognitive outcomes literature may affect the clinical decision-making process as concerns the possible risk of postsurgical cognitive morbidity, as well as conceptual inferences to be drawn regarding the role of the STN in higher-level cognitive functions. Statistical and methodological recommendations (e.g., meta-analysis) are offered to enhance the power of current and future studies examining the neuropsychological sequelae of STN DBS in PD.

INTRODUCTION

Despite myriad notable limitations to null hypothesis testing (Cohen, 1994; Donders, 2000), this traditional approach to statistical analysis remains decidedly prevalent in psychological research. As such, consideration of statistical power is critical for investigators and research consumers alike. Statistical power refers to

Address correspondence to: Steven Paul Woods, Psy.D., HIV Neurobehavioral Research Center, Department of Psychiatry (0603-H), University of California, San Diego, 150 W. Washington St., 2nd floor, San Diego, CA 92103, USA. Tel.: 619-543-5004. Fax: 619-543-1235. E-mail: spwoods@ucsd.edu
Accepted for publication: November 2, 2004.

the likelihood that one will accurately reject a null hypothesis when an effect is present (Cohen, 1988). Power is dependent on four primary factors: (a) the sample size; (b) the critical alpha level (0.05 by convention); (c) the effect size observed (or anticipated) in the population of interest; and (d) the specific statistical procedure being used. As a general rule, power values increase with larger sample sizes, stronger effects, higher critical alpha levels, and the use of tests that control more aspects of error variance. In other words, one is more likely to accurately reject a null hypothesis in a study with a large sample and liberal critical alpha level in which substantial effect sizes are evident (Hallahan & Rosenthal, 1996). The generally accepted convention for adequate power is 0.80 (range = 0, 1), which indicates that there is an 80% probability that the null hypothesis will be rejected when true effects are present (Cohen, 1992). Power values below 0.80 increase one's risk of committing a Type II error (i.e., not rejecting the null hypothesis when true population differences are present).

Cohen (1988) and a multitude of subsequent prominent investigators (e.g., Wilkinson & the American Psychological Association's Task Force on Statistical Inference, 1999) have urged behavioral scientists to perform power analyses to determine an appropriate sample size given the particular study design and hypothesized effects. Despite such longstanding recommendations and the increasing availability of resources and tools for its calculation, statistical power is not widely reported in published psychological research (e.g., Rossi, 1990; Sedlmeier & Gigerenzer, 1989). The conspicuous absence of power analyses has prompted numerous investigators over the past 30 years to conduct post hoc power analyses of specific psychological literatures. For example, systematic post hoc power reviews are available for psychotherapy (e.g., Kazantzis, 2000) and rehabilitation counseling (e.g., Kosciulek & Szymanski, 1993) outcomes, health psychology (e.g., Maddock & Rossi, 2001), and projective personality assessment (e.g., Acklin, McDowell, & Orndoff, 1992). By and large, such power reviews reach the same general conclusion: Insufficient power remains a wide spread problem in psychological research (Cohen, 1992; Sedlmeier & Gigerenzer, 1989).

In fact, the failure to consider power has been proposed as one of the "seven deadly sins" of statistical practice in clinical neuropsychology (Millis, 2003). In a recent systematic review of the neuropsychological literature, Bezeau and Graves (2001) conducted post hoc power analyses of 66 articles from the 1998 and 1999 issues of *Journal of Clinical and Experimental Neuropsychology*, *Journal of the International Neuropsychological Society*, and *Neuropsychology*. Consistent with other recent power analyses performed in the psychological literature (Maddock & Rossi, 2001; Sedlmeier & Gigerenzer, 1989), neuropsychological research generally demonstrated insufficient power to detect small and medium effect sizes. However, the median observed population effect size for the neuropsychological articles was large (Cohen's $d = 0.91$) and corresponded to an ample median power estimate of 0.93. The authors concluded that neuropsychological research typically addresses larger effect sizes than are documented in general psychological research, which may therefore allow for the use of smaller sample sizes.

Although they provide critical and informative data, statistical power reviews of broad literatures such as provided by Bezeau and Graves (2001) may not effectively generalize to specific populations and/or hypotheses (Rossi, 1990). Statistical power is particularly problematic in neuropsychological studies involving clinical populations that are difficult to recruit and enroll in research protocols. For

example, small sample sizes are endemic to studies involving persons with localized brain lesions, low base rate neurological and medical conditions (e.g., prion diseases), and/or who are undergoing novel treatment protocols (e.g., deep brain stimulation). In fact, interventional studies typically exhibit significantly lower power estimates than non-interventional studies, which is often tolerated given the novelty, potential clinical impact, and repeated-measures designs common to clinical trials (Maddock & Rossi, 2001; Vickers, 2003). Although the ethical and logistical factors underlying the smaller sample sizes in interventional studies are legitimate and difficult to circumvent, the resultant limitations on statistical power are nonetheless challenging. Neuropsychological findings derived from such small population samples are oftentimes contradictory and vary widely across published studies (e.g., Demakis, 2003), which ultimately diminishes one's ability to draw coherent clinical and conceptual inferences from the scientific literature (e.g., Cohn & Becker, 2003; Maxwell, 2004).

One notable example of this problem is evident in studies examining the cognitive effects of deep brain stimulation (DBS) of the subthalamic nucleus (STN) in persons with Parkinson's disease (PD). STN DBS is a functional neurosurgical procedure developed to reduce the cardinal motor symptoms of PD (e.g., akinesia, rigidity, and tremor) in treatment refractory patients. Briefly, the surgery involves the bilateral implantation of high-frequency stimulation, quadripolar electrodes into the STN of persons with PD. The electrodes are subsequently linked to a subcutaneous pulse generator (akin to a cardiac pacemaker) that is implanted in the subclavicular area, which allows for outpatient adjustment of stimulation parameters (i.e., frequency, pulse width, and amplitude) to maximize treatment efficacy (Rizzone et al., 2001). A growing body of literature supports the effectiveness of STN DBS for ameliorating off-motor symptoms and dyskinesias, as well as reducing antiparkinsonian medication dosages (Limousin et al., 1998; Pollak et al., 2002). The exact mechanism by which STN DBS reduces the symptoms of PD is controversial, but the high-frequency stimulation procedure may inhibit neuronal activity (e.g., membrane hyperpolarization) in the STN that, in turn, enhances the functioning of nigrostriatal motor output pathways (see Dostrovsky & Lozano, 2002, for a review).

It has been proposed that STN DBS may minimize the risk of cognitive morbidity relative to other neuroanatomical targets (e.g., globus pallidus internus) and surgical techniques (e.g., lesioning methods) (Van Horn, Schiess, & Soukup, 2001). To this end, a recent qualitative review of 16 published studies provided tentative support for the gross cognitive and neurobehavioral safety of STN DBS in PD (Woods, Fields, & Tröster, 2002). Nevertheless, the median sample size of the STN DBS studies in that review was 10 (range = 1–63, all single-group pretest–posttest designs), which raises the concern that this literature may possess inadequate power to detect significant adverse postsurgical cognitive changes. Limited statistical power—in even a small subset of studies—might also falsely increase the variability of STN DBS cognitive outcomes (i.e., adequately powered studies report adverse cognitive outcomes, whereas underpowered studies erroneously report no iatrogenic effects thereby resulting in increased variability in the literature). Indeed, inconsistencies persist across this literature regarding the extent and duration of possible changes in episodic memory, attention, and executive functions (e.g., verbal fluency) after STN DBS. For example, several investigators observed postsurgical declines on measures of verbal fluency, attention, and executive functions, perhaps mediated by the

effects of stimulation on neighboring associative and limbic fronto-striato-thalamo-cortical pathways (Woods et al., 2002). Yet the nature and extent of cognitive decrements after STN DBS is controversial as other studies report no change (and even improvement) in these same cognitive ability areas (e.g., Jahanshahi et al., 2000).

Whether STN DBS is associated with incident cognitive impairment is a question of considerable clinical (as well as conceptual) relevance. Research indicates that cognitive impairment—a common feature of PD (see Tröster & Woods, 2003 for a review)—is associated with greater difficulties independently managing one's instrumental activities of daily living (IADLs) (e.g., Cahn et al., 1998). Thus, even modest declines in executive or memory functions might be burdensome for patients who already evidence mild neuropsychological deficits prior to surgery (n.b., frank dementia is an exclusion criterion for a majority of surgical candidates). Accordingly, if limited statistical power in the STN DBS literature has masked significant postsurgical cognitive declines, such information would likely alter the informed consent process as regards the potential costs and benefits of surgery. In the absence of formal statistical power analyses, however, it is difficult to determine whether inadequate statistical power might have obscured important cognitive risks associated with STN DBS and/or contributed to inconsistent cognitive outcomes in the literature. Therefore, the aim of the present study was to provide a post hoc estimate of the statistical power of the STN DBS cognitive outcomes literature.

METHOD

To identify the relevant published articles, key search terms (e.g., subthalamic nucleus, deep brain stimulation, cognitive, etc.) were entered into the PsychINFO, PubMed, and ISI Web of Science electronic databases for the years 1997 to 2004. In addition, references from articles reporting cognitive outcomes of STN DBS were reviewed to identify other papers of interest that may not have been indexed in the aforementioned databases. To be included in the current power review, an article must have used a repeated-measures design and at least one paired-samples group-level statistical analysis (e.g., a paired-samples *t*-test) to examine the cognitive sequelae of STN DBS in a sample of persons with PD. Studies that used single- and/or mixed comparison-group designs were included. We excluded review articles, single case studies, statement papers, investigations that used only animal subjects, and studies not published in English.

The 30 articles that met study inclusion criteria were reviewed to determine whether power estimates or standardized effect sizes were reported (Alegret et al., 2001; Ardouin et al., 1999; Berney et al., 2002; Brusa et al., 2001; Burchiel, Anderson, Favre, & Hammerstad, 1999; Daniele et al., 2003; Dujardin, Defebvre, Krystkowiak, Blond, & Destee, 2001; Funkiewiez et al., 2003, 2004; Gironell, Kulisevsky, Fortuny, Garcia-Sanchez, & Pascual-Sedano, 2003; Halbig et al., 2003; Hershey et al., 2004; Hilker et al., 2003; Jahanshahi et al., 2000; Limousin et al., 1998; Lopiano et al., 2002; Moretti et al., 2003; Moro et al., 1999; Morrison et al., 2004; Patel et al., 2003; Perozzo et al., 2001; Pillon et al., 2000; Saint-Cyr, Trépanier, Rajeev, Lozano, & Lang, 2000; Schneider et al., 2003, Schroeder et al., 2003, 2004; Trépanier et al., 2000; Volkmann et al., 2001; Whelan, Murdoch, Theodoros, Hall, & Silburn, 2003; Witt et al., 2004). The *G*Power* statistical package (Buchner, Faul, &

Erdfelder, 1997; Erdfelder, Faul, & Buchner, 1996) was then used to calculate the statistical power of each study. Specifically, post hoc power calculations for paired-samples t-tests were generated considering each individual study's sample size, associated degrees of freedom, and a critical alpha level of 0.05. The effect size index $f [(\sigma_m)/\sigma]$ is recommended for study designs in which $k \geq 2$ (Cohen, 1988), as with the STN DBS literature where multiple repeated measures designs are commonplace. Accordingly, power estimates were conducted using a priori defined Cohen's f values for small ($f = 0.10$), medium ($f = 0.25$), and large ($f = 0.40$) effect sizes. Cohen's f values—which are always positive and range from zero to an indefinite upper limit—are interpreted as the standard deviation of the standardized means in a given set of populations (Cohen, 1988). Following recommendations from Zakzanis (2001) and Rossi (1990), we also calculated power estimates for very large ($f = 1.5$) Cohen's f values since traditional effect size conventions may not adequately cover the range of effects that might be of clinical interest.

In a second analysis, we derived power values specifically for verbal fluency tasks using the observed rather than a priori defined effect sizes. Verbal fluency tasks were reported in 19 (63%) of the 30 STN DBS studies, making them the most commonly employed cognitive measures in this literature. Power values were calculated for these studies using the observed effect size (f), sample size, degrees of freedom, and a critical alpha level of 0.05. We were unable to derive power values for 5 of the 19 verbal fluency studies because they did not report sufficient data to generate an effect size.

RESULTS

None of the 30 studies reported statistical power analyses or formal measures of effect size. The median sample size of persons with PD undergoing STN DBS in these studies was 14 (interquartile range = 8, 22). Descriptive statistics derived from the post hoc power analyses are presented in Table 1. Results revealed overall minimal power for the detection of conventionally small, medium, or large effect sizes (range = 0.05, 0.91). Only 7% ($n = 2$) of the studies reviewed demonstrated adequate power (≥ 0.80) to detect a traditionally large effect.

Power estimates based on observed effect sizes from the 14 studies that reported sufficient data on verbal fluency are displayed in Table 2. The mean

Table 1 Estimated power of studies reporting cognitive outcomes of STN DBS in PD ($N = 30$)

Effect size (f)	Power estimates				
	M	SD	Median	IQR	Range
Small ($f = .10$)	.07	.02	.06	.06, .07	.05, .13
Medium ($f = .25$)	.18	.13	.13	.09, .20	.06, .54
Large ($f = .40$)	.34	.23	.25	.17, .43	.07, .91
Very large ($f = 1.5$)	.94	.13	.99	.95, .99	.32, 1.00

Note. These data reflect post hoc statistical power estimates generated using standard effect size conventions (cf. observed effect sizes), which were adapted from Cohen (1988) and Zakzanis (2001). Cohen's f values [$f = (\sigma_m)/\sigma$] reflect the SD of the standardized means in a population (Cohen, 1988). DBS = deep brain stimulation; IQR = interquartile range; PD = Parkinson's disease; STN = subthalamic nucleus.

Table 2 Effect sizes and statistical power of verbal fluency changes after STN DBS in persons with PD

Verbal fluency statistic	<i>M</i>	<i>SD</i>	Median	IQR	Range
Effect size (<i>f</i>)	.23	.15	.18	.14, .30	.05, .61
Power	.16	.12	.11	.06, .23	.06, .45

Note. *N* = 14; DBS = deep brain stimulation; IQR = interquartile range; PD = Parkinson's disease; STN = subthalamic nucleus. Cohen's *f* values [$f = (\sigma_m)/\sigma$] reflect the *SD* of the standardized means in a population (Cohen, 1988).

Cohen's *f* effect size of 0.23 (*SD* = 0.15) in these studies provided a mean observed power of 0.16 (*SD* = 0.12) to detect postsurgical verbal fluency changes. Not surprisingly, the five studies that reported significant declines in verbal fluency after STN DBS demonstrated superior power (*M* = 0.25, *SD* = 0.09) to the nine that observed no such changes (*M* = 0.11, *SD* = 0.13), $X^2(1, N = 14) = 5.6$, $p = .02$, $d = 1.27$, power = 0.55.

DISCUSSION

Published studies on the neuropsychological sequelae of STN DBS in PD largely suggest that this procedure is associated with minimal risk of gross cognitive decline for a majority of appropriate surgical candidates. In support of this contention, data from the present review indicate that, on average, studies within the STN DBS literature demonstrate a 94% chance of detecting such substantial postsurgical cognitive declines (i.e., very large effect sizes) if they were truly present. However, it remains uncertain whether STN DBS leads to milder cognitive decrements in attention, verbal memory, and executive functions (see Woods et al., 2002) that nevertheless might be of clinical significance. Our review revealed surprisingly low statistical power to identify conventionally small, medium, and large effect sizes in the STN DBS cognitive outcomes literature; for example, the studies reviewed averaged only a 34% probability of accurately detecting the presence of a traditionally large effect. In fact, only two (7%) of the 30 published studies reviewed afforded sufficient power (≥ 0.80) to detect a hypothesized large effect size. Low power was evident even when we examined the observed medium effect sizes associated with postsurgical changes in verbal fluency, which was the most commonly assessed domain. Notably, studies that reported significant postsurgical verbal fluency declines displayed superior power to those that observed no effect of DBS on verbal fluency performance.

It is widely held that the substantial gains in motor functioning and health-related quality of life after STN DBS (e.g., Pollak et al., 2002) outweigh the risk of cognitive decline for a large proportion of surgical candidates (Woods et al., 2002). However, evidence for low statistical power to detect small, medium, and large effect sizes precludes one from drawing conclusions regarding the full impact of STN DBS on cognitive functions. This is of considerable importance because Type II error is especially risky when assessing cognitive morbidity associated with STN DBS (cf., an elevated Type I error risk would fall conservatively in the direction of safety). While the presence of Type II error in interventional studies designed to detect the benefits of a given procedure may result in the erroneous

conclusion that a given treatment is ineffective (Maddock & Rossi, 2001), false negatives in the detection of adverse side effects are potentially more perilous. Indeed, postsurgical cognitive decrements associated with large (and perhaps even medium) effect sizes may adversely impact performance of IADLs for persons with PD (e.g., Cahn et al., 1998; Chaytor & Schmitter-Edgecombe, 2003), especially for patients with mild presurgical cognitive deficits for whom even a slight decrement in neuropsychological performance may lead to IADL complications. Accordingly, the possibility of significant Type II error in the existing STN DBS cognitive outcomes literature might influence the clinical decision-making process regarding the risk–benefit ratio of cognitive morbidity and considerable motor gains associated with this procedure. Surgical candidates and their caregivers should be informed regarding the possible risk of unforeseen cognitive decrements associated with STN DBS. A postsurgical neuropsychological evaluation may be indicated to assess the possible incidence of subtle cognitive, psychiatric, and/or functional impairment, as well as to inform interventions that would maximize adherence to postsurgical medical regimens (Woods et al., 2002).

Inadequate statistical power necessitates cautious interpretation of the conceptually driven investigations of the STN's involvement in higher-level cognitive functions. Given the relative ease with which stimulation parameters may be manipulated on an outpatient basis, DBS provides the cognitive neuropsychologist a unique opportunity to employ more rigorous, hypothesis-driven experimental methodologies. In response, emerging studies are exploring the nature and extent of the STN's role in specific aspects of language, executive functions, and social cognition using dissociation methodologies (e.g., on-off-on stimulation designs) that require acceptance of a true null hypothesis. Nevertheless, absence of evidence cannot be taken as convincing evidence of absence when interpreting the literature regarding the neuropsychological sequelae of STN DBS in persons with PD. As eloquently stated by Cohen (1988):

An analysis which finds that the power was low should lead one to regard the negative results as ambiguous, since failure to reject the null hypothesis cannot have much substantive meaning when, even though the phenomenon exists (to some given degree), the a priori probability of rejecting the null hypothesis was low. (p. 4)

The small samples in the STN DBS literature are ostensibly a function of logistical and ethical problems inherent to research evaluating novel neurosurgical procedures (see Fields & Tröster, 2000). Although the use of repeated-measures methodologies may increase study power by reducing variability parameters (Vickers, 2003), investigators are nevertheless encouraged to maximize the number of enrolled study participants. Such efforts will likely be facilitated by the increasing availability of STN DBS subsequent to its approval by the Food and Drug Administration. Ideally, sample sizes would be dictated by a priori power analyses. Numerous texts (e.g., Cohen, 1988), published articles (e.g., Hallahan & Rosenthal, 1996), and computer software packages (e.g., G*Power: Buchner et al., 1997) are readily accessible in this regard. Sample sizes informed by a priori power analyses will increase statistical rigor, as well as afford investigators the opportunity to utilize more complex statistical procedures to examine possible mediators of postsurgical

cognitive changes (e.g., age, presurgical cognitive deficits, psychiatric illness, stimulation parameters).

A few limitations of the current study should be highlighted. Firstly, not all of the studies included in this review were designed for the primary purpose of evaluating cognitive outcomes. Secondly, the post hoc statistical power analyses reported herein were conducted specifically for paired-samples statistical tests and therefore do not necessarily generalize to other reported statistical analyses (e.g., between-group comparisons or regression-based analyses) (see Rossi, 1990). Thirdly, the apparent variability in the STN DBS cognitive outcomes literature may be partly attributable to factors other than low statistical power. For instance, heterogeneity in participant demographics and disease characteristics, surgical techniques, stimulation parameters, variable test-retest intervals, practice effects, postsurgical medication changes, and/or Type I error due to multiple exploratory statistical comparisons (see Maxwell, 2004) might also contribute to inconsistent findings (see Woods et al., 2002, for review). When appropriate, investigators might therefore consider decreasing sample heterogeneity, using highly reliable and valid dependent measures with continuous outcome variables (cf., dichotomous dependent variables), limiting critical alpha corrections, and pooling data across multiple research centers in an effort to increase power (e.g., Hallahan & Rosenthal, 1996; Maddock & Rossi, 2001).

Finally, meta-analyses are another means of potentially increasing the statistical power of existing literatures that, like STN DBS, are hampered with small sample sizes (e.g., Demakis, 2003). Enhanced statistical power is one of the most commonly cited benefits of meta-analytic studies (Cohn & Becker, 2003). A fundamental aim of a meta-analysis is to estimate a population effect size (θ) by examining findings across independent studies (Demakis, 2006). Meta-analyses can increase statistical power by lowering the standard error associated with the population effect size, which ultimately provides a smaller confidence interval and thereby increases one's power to detect true nonzero population effects (Cohn & Becker, 2003). Meta-analyses would also allow for a more precise and powerful examination of potential moderator variables (e.g., stimulation parameters) that might influence the neuropsychological outcomes of STN DBS. A priori power analyses should also be considered to evaluate the risk of Type II error for meta-analyses, particularly when studies with small sample sizes are involved (Hedges & Pigott, 2001).

ACKNOWLEDGEMENT

The authors thank Sarah Gibson for her assistance in compiling the references.

REFERENCES

- Acklin, M. W., McDowell, C. J., & Orndoff, S. (1992). Statistical power and the Rorschach: 1975–1991. *Journal of Personality Assessment*, *59*, 366–379.
- Alegret, M., Junque, C., Valldeoriola, F., Vendrell, P., Pilleri, M., Rumia, J., et al. (2001). Effects of bilateral subthalamic stimulation on cognitive function in Parkinson disease. *Neurology*, *58*, 1223–1227.

- Ardouin, C., Pillon, B., Peiffer, E., Bejjani, P., Limousin, P., Damier, P., et al. (1999). Bilateral subthalamic and pallidal stimulation for Parkinson's disease affects neither memory nor executive functions: A consecutive series of 62 patients. *Annals of Neurology*, *46*, 217–223.
- Berney, A., Vingerhoets, F., Perrin, A., Guex, P., Villemure, J., Burkhard, P., et al. (2002). Effect on mood of subthalamic DBS for Parkinson's disease. *Neurology*, *59*, 1427–1429.
- Bezeau, S., & Graves, R. (2001). Statistical power and effect sizes of clinical neuropsychology research. *Journal of Clinical and Experimental Neuropsychology*, *23*, 399–406.
- Brusa, L., Pierantozzi, M., Peppe, A., Altibrandi, M., Giacomini, P., & Mazzone, P. (2001). Deep brain stimulation (DBS) attentional effects parallel those of l-dopa treatment. *Journal of Neural Transmission*, *108*, 1021–1027.
- Buchner, A., Faul, F., & Erdfelder, E. (1997). *G*Power: A priori, post hoc, and compromise power analyses for the Macintosh* (Version 2.1.2) [Computer program]. Trier, Germany.
- Burchiel, K., Anderson, V., Favre, J., & Hammerstad, J. (1999). Comparison of pallidal and subthalamic nucleus deep brain stimulation for advanced Parkinson's disease: Results of a randomized, blinded pilot study. *Neurosurgery*, *45*, 1375–1384.
- Cahn, D. A., Sullivan, E. V., Shear, P. K., Pfefferbaum, A., Heit, G., & Silverberg, G. (1998). Differential contributions of cognitive and motor component processes to physical and instrumental activities of daily living in Parkinson's disease. *Archives of Clinical Neuropsychology*, *13*, 575–583.
- Chaytor, N. & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*, 181–197.
- Cohen, J. (1988). *Statistical power analyses for the behavioral sciences* (2nd edition). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997–1003.
- Cohn, L. D. & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods*, *8*, 243–253.
- Daniele, A., Albanese, A., Contarino, M., Zinzi, P., Barbier, A., Gasparini, F., et al. (2003). Cognitive and behavioral effects of chronic stimulation of the subthalamic nucleus in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *74*, 175–182.
- Demakis, G. (2003). A meta-analytic review of the sensitivity of the Wisconsin Card Sorting Test to frontal and lateralized frontal brain damage. *Neuropsychology*, *17*, 255–264.
- Demakis, G. (2006). Meta-analysis in neuropsychology: Basic approaches, findings, and applications. *The Clinical Neuropsychologist*.
- Donders, J. (2000). From null hypothesis to clinical significance. *Journal of Clinical and Experimental Neuropsychology*, *22*, 265–266.
- Dostrovsky, J. O. & Lozano, A. M. (2002). Mechanisms of deep brain stimulation. *Movement Disorders*, *17*(Suppl. 3), S63–S68.
- Dujardin, K., Defebvre, L., Krystkowiak, P., Blond, S., & Destee, A. (2001). Influence of chronic bilateral stimulation of the subthalamic nucleus on cognitive function in Parkinson's disease. *Journal of Neurology*, *248*, 603–611.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). G*Power: A general power analysis program. *Behavior Research Methods, Instruments, and Computers*, *28*, 1–11.
- Fields, J. A. & Tröster, A. I. (2000). Cognitive outcomes after deep brain stimulation for Parkinson's disease: A review of initial studies and recommendations for future research. *Brain and Cognition*, *42*, 268–293.
- Funkiewiez, A., Ardouin, C., Caputo, E., Krack, P., Fraix, V., Klinger, H., et al. (2004). Long term effects of bilateral subthalamic nucleus stimulation on cognitive function, mood,

- and behaviour in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, 75, 834–839.
- Funkiewiez, A., Ardouin, C., Krack, P., Fraix, V., Van Blercon, N., Xie, J., et al. (2003). Acute psychotropic effects of bilateral subthalamic nucleus stimulation and levodopa in Parkinson's disease. *Movement Disorders*, 18, 524–530.
- Gironell, A., Kulisevsky, J., Fortuny, N., Garcia-Sanchez, C., & Pascual-Sedano, B. (2003). Effects of pallidotomy and bilateral subthalamic stimulation on cognitive function in Parkinson's disease. *Journal of Neurology*, 250, 917–923.
- Halbig, T., Gruber, D., Kopp, U., Scherer, P., Schneider, G., Trottenberg, T., et al. (2003). Subthalamic stimulation differentially modulates declarative and nondeclarative memory. *NeuroReport*, 15, 539–543.
- Hallahan, M. & Rosenthal, R. (1996). Statistical power: Concepts, procedures, and applications. *Behavior Research and Therapy*, 34, 489–499.
- Hedges, L. V. & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, 6, 203–217.
- Hershey, T., Revilla, F., Wernle, A., Gibson, P., Dowling, J., & Perlmutter, J. (2004). Stimulation of STN impairs aspects of cognitive control in PD. *Neurology*, 62, 1110–1114.
- Hilker, R., Voges, J., Weisenbach, S., Kalbe, E., Burghaus, L., Ghaemi, M., et al. (2003). Subthalamic nucleus stimulation restores glucose metabolism in associative and limbic cortices and in cerebellum: evidence from a FDG-PET study in advanced Parkinson's disease. *Journal of Cerebral Blood Flow and Metabolism*, 24, 7–16.
- Jahanshahi, M., Ardouin, C. M. A., Brown, R. G., Rothwell, J. C., Obeso, J., Albanese, A., et al. (2000). The impact of deep brain stimulation on executive function in Parkinson's disease. *Brain*, 123, 1142–1154.
- Kazantzis, N. (2000). Power to detect homework effects in psychotherapy outcome research. *Journal of Consulting and Clinical Psychology*, 68, 166–170.
- Kosciulek, J. F. & Szymanski, E. M. (1993). Statistical power analysis of rehabilitation counseling research. *Rehabilitation Counseling Bulletin*, 36, 212–219.
- Limousin, P., Krack, P., Pollak, P., Benazzouz, A., Ardouin, C., Hoffmann, D., et al. (1998). Electrical stimulation of the subthalamic nucleus in advanced Parkinson's disease. *New England Journal of Medicine*, 339, 1105–1111.
- Lopiano, L., Rizzone, M., Bergamasco, B., Tavezza, A., Torre, E., Perozzo, P., et al. (2002). Deep brain stimulation of the subthalamic nucleus in PD: An analysis of the exclusion causes. *Journal of Neurological Sciences*, 195, 167–170.
- Maddock, J. E. & Rossi, J. S. (2001). Statistical power of articles published in three health psychology-related journals. *Health Psychology*, 20, 76–78.
- Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, 9, 147–163.
- Millis, S. R. (2003). Statistical practices: The seven deadly sins. *Child Neuropsychology*, 9, 221–233.
- Moretti, R., Torre, P., Antonello, R. M., Capus, L., Marsala, S., Cattaruzza, T., et al. (2003). Neuropsychological changes after subthalamic nucleus stimulation: A 12 month follow-up in nine patients with Parkinson's disease. *Parkinsonism and Related Disorders*, 10, 73–79.
- Moro, E., Scerrati, M., Romito, L. M. A., Roselli, R., Tonali, P., & Albanese, A. (1999). Chronic subthalamic nucleus stimulation reduces medication requirements in Parkinson's disease. *Neurology*, 53, 85–90.
- Morrison, C. E., Borod, J. C., Perrine, K., Beric, A., Brin, M. F., Rezaei, A., et al. (2004). Neuropsychological functioning following bilateral subthalamic nucleus stimulation in Parkinson's disease. *Archives of Clinical Neuropsychology*, 19, 165–181.

- Patel, N. K., Plaha, P., O'Sullivan, K., McCarter, R., Heywood, P., & Gill, S. S. (2003). MRI directed bilateral stimulation of the subthalamic nucleus in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *74*, 1621–1637.
- Perozzo, P., Rizzone, M., Bergamasco, B., Castelli, L., Lanotte, M., Tavella, A., et al. (2001). Deep brain stimulation of the subthalamic nucleus in Parkinson's disease: Comparison of pre- and postoperative neuropsychological evaluation. *Journal of the Neurological Sciences*, *192*, 9–15.
- Pillon, B., Ardouin, C., Damier, P., Krack, P., Houeto, J. L., Klinger, H., et al. (2000). Neuropsychological changes between “off” and “on” STN or GPi stimulation in Parkinson's disease. *Neurology*, *55*, 411–418.
- Pollak, P., Fraix, V., Krack, P., Moro, E., Mendes, A., Chabardes, S., et al. (2002). Treatment results: Parkinson's disease. *Movement Disorders*, *17*(Suppl. 3), S75–S83.
- Rizzone, M., Lanotte, M., Bergamasco, B., Tavella, A., Torre, E., Faccani, G., et al. (2001). Deep brain stimulation of the subthalamic nucleus in Parkinson's disease: Effects of variation in stimulation parameters. *Journal of Neurology, Neurosurgery, and Psychiatry*, *71*, 215–219.
- Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? *Journal of Consulting and Clinical Psychology*, *58*, 646–656.
- Saint-Cyr, J. A., Trépanier, L. L., Rajeev, K., Lozano, A., & Lang, A. E. (2000). Neuropsychological consequences of chronic bilateral stimulation of the subthalamic nucleus in Parkinson's disease. *Brain*, *123*, 2091–2108.
- Schneider, F., Habel, U., Volkmann, J., Regel, S., Kornischka, J., Sturm, V., et al. (2003). Deep brain stimulation of the subthalamic nucleus enhances emotional processing in Parkinson's disease. *Archives of General Psychiatry*, *60*, 296–302.
- Schroeder, U., Kuehler, A., Hennenlotter, A., Haslinger, B., Tronnier, V. M., Krause, M., et al. (2004). Facial expression recognition and subthalamic nucleus stimulation. *Journal of Neurology, Neurosurgery, and Psychiatry*, *75*, 648–650.
- Schroeder, U., Kuehler, A., Lange, K. W., Haslinger, B., Tronnier, V. M., Krause, M., et al. (2003). Subthalamic nucleus stimulation affects a frontotemporal network: a PET study. *Annals of Neurology*, *54*, 445–450.
- Sedlmeier, P. & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*, 309–316.
- Trépanier, L. L., Kumar, R., Lozano, A. M., Lang, A. E., & Saint-Cyr, J. A. (2000). Neuropsychological outcome of GPi pallidotomy and GPi or STN deep brain stimulation in Parkinson's disease. *Brain and Cognition*, *42*, 324–347.
- Tröster, A. I. & Woods, S. P. (2003). Neuropsychological aspects of Parkinson's disease and parkinsonian syndromes. In R. Pahwa, K. Lyons, & W.C. Koller (Eds.), *Handbook of Parkinson's disease* (3rd edition) (pp. 127–157). New York: Marcel Dekker.
- Van Horn, G., Schiess, M. C., & Soukup, V. M. (2001). Subthalamic deep brain stimulation: Neurobehavioral concerns. *Archives of Neurology*, *58*, 1205–1206.
- Vickers, A. J. (2003). How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC Medical Research Methodology*, *3*, 22.
- Volkmann, J., Allert, N., Voges, J., Weiss, P. H., Freund, H. J., & Sturm, V. (2001). Safety and efficacy of pallidal or subthalamic nucleus stimulation in advanced PD. *Neurology*, *56*, 548–551.
- Whelan, B. M., Murdoch, B. E., Theodoros, D. G., Hall, B., & Silburn, P. (2003). Defining a role for the subthalamic nucleus within operative theoretical models of subcortical participation in language. *Journal of Neurology, Neurosurgery, and Psychiatry*, *74*, 1543–1550.
- Wilkinson, L. & The American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594–604.

- Witt, K., Pulkowski, U., Herzog, J., Lorenz, D., Hamel, W., Deuschl, G., et al. (2004). Deep brain stimulation of the subthalamic nucleus improves cognitive flexibility but impairs response inhibition in Parkinson's disease. *Archives of Neurology*, *61*, 697–700.
- Woods, S. P., Fields, J. A., & Tröster, A. I. (2002). Neuropsychological sequelae of subthalamic nucleus deep brain stimulation in Parkinson's disease: A critical review. *Neuropsychology Review*, *12*, 111–126.
- Zakzanis, K. K. (2001). Statistics to tell the truth, the whole truth, and nothing but the truth: Formulae, illustrative numerical examples, and heuristic interpretation of effect size analyses for neuropsychological researchers. *Archives of Clinical Neuropsychology*, *16*, 653–667.