Pergamon

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Addictive Behaviors 29 (2004) 673–684

ADDICTIVE BEHAVIORS

The blind spot in the nicotine replacement therapy literature: Assessment of the double-blind in clinical trials

Marc Mooney*, Thom White [1], Dorothy Hatsukami

Tobacco Use Research Center, University of Minnesota, 2701 University Avenue S.E., Suite 201, Minneapolis, MN 55414, USA

Abstract

While clinical trials of medications often use a double-blind procedure, the integrity of the blind and its relationship to treatment outcome is seldom examined. In this review, 73 double-blind, placebo-controlled clinical trials of the nicotine replacement therapies (NRTs) in smoking cessation were identified. Seventeen articles were found that assessed blindness integrity, demonstrating major variations in the assessment, analysis, and reporting of blindness integrity. Although 12 studies found that subjects accurately judged treatment assignment at a rate significantly above chance, the available literature does not permit definitive conclusions about blindness integrity. Recommendations for the assessment, analysis, and reporting of blindness integrity are made.

Keywords: Double blindness; Placebo effect; Nicotine replacement; Compliance; Review

1. Introduction

Systematic error (bias) in research studies, such as clinical trials of pharmacological therapies, can prevent the valid estimation of treatment effects (Friedman, Furburg, & DeMets, 1995). Expectations by research participants and experimenters about a treatment can be a significant source of bias. The double-blind, placebo-controlled trial has been an essential bulwark against these expectancy biases (Friedman et al., 1995; Nash, 1957).

* Corresponding author. Tel.: +1-713-500-2810; fax: +1-713-500-2849.
E-mail address: Marc.E.Mooney@uth.tmc.edu (M. Mooney).
[1] Thom White is now at the Department of Psychology, University of Maryland.

0306-4603/$ – see front matter
doi:10.1016/j.addbeh.2004.02.010

Ideally, investigators and participants are kept blind to drug assignment (i.e., active vs. placebo), and biases toward or against a specific treatment cannot taint the results of the study. However, it is generally believed that study blindness is often far from perfect (Fisher & Greenberg, 1993; Greenberg, Bornstein, Greenberg, & Fisher, 1992; Margraf et al., 1991; Oxtoby, Jones, & Robinson, 1989). Several mechanisms can lead to study "unblinding," including the presence or absence of side effects or treatment effects, which experimenters and participants use to determine treatment assignment.

One class of pharmacotherapy, the nicotine replacement therapies (NRTs), may be particularly vulnerable to study unblinding. First, nicotine is a psychoactive drug and, by definition, more easy to discriminate than a nonpsychoactive drug (e.g., a placebo) due to the interoceptive cues it provides (Perkins et al., 1996). Smokers and nicotine-naive participants can be trained to reliably distinguish various doses of nicotine from placebo (Perkins et al., 1996; Perkins, DiMarco, Grobe, Scierka, & Stiller, 1994; Perkins, Sanders, D'Amico, & Wilson, 1997), although training and rate of delivery influence sensitivity of discrimination. Second, NRTs have been shown to reliably reduce withdrawal symptoms and craving (Hughes et al., 1984). Smokers vary in severity of withdrawal and craving, but most have quit several times before entering clinical trials, and many are only too familiar with the syndrome (Hughes, Gust, & Pechacek, 1987). The use of a double-blind design provides no guarantee that experimenters and participants remain blind to their treatment assignment.

Given these and other potential threats to blindness, Hughes and Krahn (1985) have described a multistage procedure for assessing whether blindness was maintained and, if not, whether it influenced treatment effects. The first step is determining whether participants or experimenters can judge treatment assignment better than chance (i.e., whether *blindness failure* has occurred). If blindness failure occurs, then a second test is conducted to determine if the ability to accurately guess treatment assignment affected treatment results (i.e., whether *blindness bias* has occurred). If blindness bias occurs, statistical procedures controlling for its effects (*bias adjustment*) must be applied to maintain the quality of the results. The goal of the current paper is to determine how closely double-blind, placebo-controlled NRT trials have followed the recommendations of this methodological study (Hughes & Krahn, 1985). Therefore, the current review has three aims. First, the review will describe the methods of assessing and reporting blindness testing in the nicotine gum, patch, nasal spray, and inhaler literatures. Second, the frequency of blindness failure and blindness bias will be reported. Third, the frequency and effect of bias adjustment will be presented.

2. Method

2.1. Review design

2.1.1. Study selection

Initial selection of studies for the present review began with a search of three widely used computer databases, PsychINFO (American Psychological Association), MEDLINE (National Library of Medicine), and Digital Dissertations (Bell Howell). Queries were

conducted employing combinations of the key words double blindness, veiling, smoking, smoking cessation, nicotine, nicotine replacement therapy, nicotine gum, nicotine patch, nicotine spray, and nicotine inhaler. In addition, a request was made through the Society for Research on Nicotine and Tobacco (an international scientific organization) LISTSERV for unpublished data from NRT clinical trials on tests of participant and experimenter blindness. Reference lists in the collected literature were examined for potentially appropriate studies. The collection of studies ended January 1, 2002.

2.1.2. Study inclusion and exclusion criteria

Studies included in the present review were those that (a) employed nicotine gum, the nicotine patch, the nicotine nasal spray, or the nicotine inhaler, (b) were double blind, (c) included a placebo control, (d) were clinical trials in which the primary outcome measure was proportion of smokers who quit, (e) were written in English or had been translated into English, and (f) were presented in manuscript format. Among the studies that met these inclusion criteria, studies were then selected that reported data on blindness integrity (i.e., participant and/or experimenter judgments and analyses of these data). Conference abstracts and FDA drug applications were not included in the current review.

2.1.3. Variables coded

Data were entered and prepared for analysis using Access 97 (Microsoft Access, 1997) and Excel 97 (Microsoft Excel, 1997). Two categories of variables were selected for consideration in this paper. First, variables relating to the methods of blindness assessment were coded as follows: (a) number of studies assessing blindness; (b) type of judge guessing treatment condition (i.e., participant, experimenter, or both); (c) number of judges; (d) time of judgment after drug treatment begins; (e) rationale for judgment (e.g., side effects); (f) judges' confidence in judgment; (g) whether statistical tests of blindness failure were conducted; (h) whether statistical tests of blindness bias (Drug Assignment × Drug Judgment interaction) were conducted; and (i) whether statistical adjustments of blindness bias were conducted. Second, the following outcome variables were coded: (j) proportions of correct, incorrect, and uncertain judges; (k) numbers of correct, incorrect, and uncertain judges; (l) outcome of test of blindness failure; (m) outcome of test of influence of blindness bias; (n) quit rates adjusted for blindness failure; and (o) quit rates unadjusted for blindness failure. In the case of item (l), blindness failure was defined as a significant chi-square test of independence, while for item (m), blindness bias was determined by a significant between drug assignment and drug judgment interaction (e.g., a logistic regression). All judgment percentages are based on participant reports because few studies ($n = 4$) reported experimenter judge data.

2.1.4. Interrater agreement

Chance-corrected measures of interrater agreement between two of the authors (M. M. and T. W.) (Orwin, 1994) were computed for the 14 variables coded (dichotomous measures, Cohen's kappa, κ; polychotomous and continuous measures, intraclass correlation [ICC]). Correlations were computed using the FREQ procedure and the INTRAC macro in SAS

Version 8.01 (SAS, 2002). For κ, values between 0.40 and 0.59 are considered fair, those between 0.60 and 0.74 are considered good, and those above 0.74 are considered excellent (Orwin, 1994). For noncontinuous variables ($n = 7$), κs ranged from 0.94 to 1.0 (median = 0.99). ICCs ranged ($n = 7$) from 0.47 to 1.00 (median = 0.69). Accordingly, the majority of variables were coded with good or excellent reliability. Typical discrepancies involved computational or clerical errors and were resolved by discussion.

2.2. Analyses

Analyses were conducted using SAS Version 8.01 (SAS, 2002). Descriptive statistics were generated using FREQ and UNIVARIATE procedures. Random effects meta-analyses of judgment data for a small number of available nicotine patch studies (the only literature populous enough to allow meta-analysis) were conducted to estimate the overall statistical significance of blindness failure using PROC FREQ (SAS, 2002). Each meta-analysis was conducted by combining the number of individuals judging correctly or incorrectly within active and placebo conditions across available studies, extending the single-study analysis described by Hughes and Krahn (1985). Several nicotine patch studies excluded judgment data from unconfident participants or conducted tests of blindness with or without these data (e.g., see Westman, Levin, & Rose, 1993). Meta-analyses were computed separately for studies including or excluding unconfident judges.

3. Results

3.1. Assessment methodology

3.1.1. Assessment frequency

Methods used to assess study blindness are displayed by route of administration (see Table 1). Relatively few nicotine gum ($n = 2$, $N_{\text{Total}} = 26$) and nicotine patch ($n = 10$, $N_{\text{Total}} = 39$) studies provided any information on testing blindness integrity, while about half of the nicotine spray ($n = 2$, $N_{\text{Total}} = 4$) studies did. The majority of nicotine inhaler studies ($n = 3$, $N_{\text{Total}} = 4$) provided information on blindness integrity testing.[2]

3.1.2. Rater

All identified studies that assessed for blindness asked participants to judge their drug assignment, while only four reported asking experimenters (nurses, Anonymous, 1993; training personnel, Buchkremer, Bents, Horstmann, Opitz, & Tolle, 1989; therapists, Hall, Tunstall, Ginsberg, Benowitz, & Jones, 1987; physicians and pharmacists, Hughes, Gust, Keenan, Fenwick, & Healey, 1989). In addition, only one study that queried experimenters reported any specific analyses (i.e., Hall et al., 1987).

[2] Only those studies providing blindness data are referenced. Please write the first author for a complete list of these studies.

Table 1
Methodology of blindness assessment in double-blind, placebo-controlled NRTs

	Gum	Patch	Spray	Inhaler	Total
Sample size					
N_{Tested}[a]	2	10	2	3	17
N_{Total}[b]	26	39	3	4	73
Rating					
Time[c]	1	7	1	3	12
Median time[d]	12	12	52	52	19
No. rating[e]	1	6	1	1	9
Judge					
Participant	2	10	2	3	17
Experimenter	2	2	0	0	4
Rationale[f]	0	1	0	0	1
Confidence[g]	1	4	0	1	6

[a] Number of studies that assessed blindness integrity.
[b] Number of double-blind, placebo-controlled located.
[c] Number of studies reporting time of judgment.
[d] Median time of judgment in weeks.
[e] Number of studies reporting the number of participants at judgment.
[f] Reasons for judgment collected.
[g] Confidence in judgment determined through separating those confident to judge and those unconfident to do so.

3.1.3. Time and sample size at judgment

Of the 17 identified studies, 12 (71%) clearly reported the time at which blindness was assessed. Considerable variation in time of assessment was observed (median = 19 weeks, range = 6–52 weeks). Eighty-three percent of the studies assessed blindness during or at the end of NRT treatment. Slightly more than half of the studies ($P = 56\%$) reported the number of participants who judged their drug assignment, while 16 studies reported at least percentages (exception, Buchkremer et al., 1989).

3.1.4. Rationale and confidence

Only one of the identified studies reported asking participants about their reasons for judgment (Abelin, Buehler, Muller, Vesanen, & Imhof, 1989). More than a third of studies required participants that were too uncertain to judge their drug condition to abstain from rating. No studies reported asking participants to rate confidence on a separate scale.

3.2. Blindness failure

3.2.1. Frequency of failure

Twelve of the 17 studies (71%) reported a test of blindness failure ($n = 2$, nicotine gum; $n = 6$, nicotine patch; $n = 7$, nicotine nasal spray; and $n = 1$, nicotine inhaler). Importantly, 6 studies, although reporting rates of accurate and inaccurate judgments, present no results of a test of statistical significance (Abelin et al., 1989; Gourlay, Forbes, Marriner, Pethica, &

Table 2
Meta-analysis of five nicotine patch studies including unconfident judges

	Correct N (%)[a]	Incorrect N (%)
Placebo	531 (52.4)	482 (47.6)
Active[b]	618 (59.1)	427 (40.9)

A significant effect for blindness failure was found [$\chi^2(1, N=1968)=9.82$, $P<.005$].

[a] Percentages are computed within each drug assignment.

[b] The proportion of those receiving active treatment guessing correctly was significantly greater than those guessing incorrectly [$\chi^2(1, N=1001)=32.0$, $P<.001$].

McNeil, 1995; Hjalmarson, Nilsson, Sjostrom, & Wiklund, 1997; Lin, 1996; Schneider et al., 1995; Tonnesen, Norregaard, Mikkelsen, Jorgensen, & Nilsson, 1993). In the 6 studies that did not report blindness tests, chi-square tests of independence were conducted (Hughes & Krahn, 1985); 2 of the 6 studies showed blindness failures (Gourlay et al., 1995; Schneider et al., 1995).

3.2.2. Meta-analysis of failure

In an effort to estimate the statistical significance of blindness failure within a route of administration, two meta-analyses using nine nicotine patch studies[3] were conducted for papers including uncertain judges (Abelin et al., 1989; Anonymous, 1993; Gourlay et al., 1995; Killen, Fortmann, Davis, & Varady, 1997; Tonnesen, Norregaard, Simonsen, & Sawe, 1991) or excluding uncertain judges (Ahluwalia, McNagny & Clark, 1998; Lin, 1996; Sonderskov, Olsen, Sabroe, Meillier, & Overvad, 1997; Westman et al., 1993). The studies of the other NRT routes (gum, inhaler, nasal spray) identified were too few to allow meta-analysis.

The meta-analysis of the five patch studies that included unconfident judges found that participants were able to judge their drug assignment at rate better than chance [$\chi^2(1, N=1968)=9.82$, $P<.005$], and that this unblinding owed specifically to those in the active patch condition being able to identify their drug assignment [$\chi^2(1, N=1001)=32.0$, $P<.001$] (Table 2).

In the meta-analysis of the four patch studies that excluded unconfident judges, a statistical trend toward blindness failure was observed [$\chi^2(1, N=798)=3.1$, $P<.10$]. Participants who did not judge their drug assignment due to the lack of confidence were equally divided between the active ($P=20\%$) and placebo groups ($P=20\%$) (Table 3).

3.2.3. Blindness bias and adjustment

Despite the fact that 12 studies reported a blindness failure, only three gave some indication of testing for blindness bias (Hall et al., 1987; Hughes et al., 1989; Tonnesen et al., 1991). None of these three studies found a significant interaction between drug

[3] Buchkremer et al. (1989) reported no statistically significant blindness failure, but failed to report sample sizes or proportions judging correctly or incorrectly. Accordingly, this study was not included in the meta-analysis.

Table 3
Meta-analysis of four nicotine patch studies excluding unconfident judges

	Correct *N* (%)[a]	Incorrect *N* (%)
Placebo	253 (63.6)	145 (36.4)
Active	230 (57.5)	170 (42.5)

In a second meta-analysis of four patch studies excluding unconfident judges, a statistical trend toward blindness failure was observed [$\chi^2(1, N=798)=3.1$, $P<.10$].

[a] Percentages are computed within each drug assignment.

assignment and drug judgment requiring for a statistical or analytical correction for blindness bias.

4. Discussion

A survey of 73 double-blind, placebo-controlled nicotine replacement trials identified just 17 studies that made some effort to assess the integrity of their double-blind procedures. These few studies provide insufficient evidence for definitive conclusions about the overall integrity of blindness in the NRT literature, but provide the basis for several significant recommendations.

Overall, relatively few clinical trials of NRT have reported assessing for blindness failure, and these studies have provided limited information about assessment methods. Studies reporting on blindness integrity assessed for blindness at widely variable times, often failing to indicate sample sizes at the time of judgment. Experimenter judges were seldom asked to provide ratings. Information about reasons for judgments and confidence in judgments were rarely elicited. Over half of the 17 studies found a blindness failure, and meta-analyses including a small sample of potentially unrepresentative nicotine patch studies also suggest a trend for blindness failure. Beyond the initial question of blindness failure, even less can be said concerning the frequency of blindness bias, its magnitude, and blindness bias adjustment. Only three studies, none of which involved the nicotine patch, reported testing for blindness bias, with none finding any bias. Overall, the NRT literature has been largely silent on blindness failure, blindness bias, or blindness bias adjustment. The limited amount of information reported has been assessed and reported with inconsistency.

4.1. Limitations

One limitation of this paper is the limited number of studies analyzed and possibility of an *availability bias* in this sample. Availability bias refers to the assertion that studies available for review (i.e., published studies) might have higher effect sizes than do unpublished studies not available for review (Hunter & Schmidt, 1990). The authors attempted to locate as many published and unpublished reports as possible assessing the integrity of study blindness. Nevertheless, it is possible that studies meeting inclusion criteria were not identified. In addition, it is possible that identified studies that did not report on blindness failure actually

tested blindness integrity, found it to be intact, and neglected to report this in the manuscript. A second potential shortcoming relates to the use of meta-analysis in this paper. Effect sizes taken from different literatures should not be combined to avoid the so-called the "apples and oranges" effect (Hunter & Schmidt, 1990). In the current review, the outcome of interest was judgment of treatment assignment and thus could be combined readily across studies (unlike other meta-analyses that combine theoretically homogenous outcomes that must first be converted to a common metric, such as Cohen's *d*). Furthermore, meta-analyses were conducted only within one route of administration, the nicotine patch, and were further separated into those that included or excluded unconfident judges.

4.2. Future directions

The NRT literature could be greatly strengthened by the adoption of several points concerning testing for blindness failure, analysis of blindness failure, and experimental design.

4.2.1. Assessment methods

Almost all studies reported assessing blindness only once (for an exception, see Hall et al., 1987). The bases for the conclusions made by judges may change depending on the stage of the study (Margraf et al., 1991), so blindness should be assessed at different points, and studies need to clearly report when the assessments are made. Although repeated questioning may "sensitize" participants and experimenters into contemplating the blind, the effects of the number and frequency of questionings on the accuracy of guessing is readily testable. By testing the blind in both participants and experimenters over time, the role of each group potentially breaking the blind and biasing the other can be ascertained. Both participants and experimenters should serve as judges in double-blind trials to confirm the symmetry of the blind and to indicate which group, if any, first became "unblinded."

Second, judges should be questioned about their reasons for and confidence in judgment to help identify weak spots of the blind and mask them better in the future. Confidence should be assessed on a common metric (e.g., 7-point Likert scale or a visual analog scale) to facilitate cross-study review and more sophisticated analysis of this variable.

4.2.2. Analysis and reporting

It is insufficient for NRT studies to show that blindness did not hold (Hughes & Krahn, 1985). When unblinding is found to have occurred in a study, the experimenters must test for an interaction between drug assignment and drug judgment for the dependent measures. Moreover, if this interaction is significant, experimenters should attempt to provide an estimate of treatment effect that is not biased by blindness failure.

NRT trials also need to clearly report how many participants act as judges at each time a judgment is made because this number tends to decline throughout the course of treatment (Hall et al., 2001). In particular, differential attrition, in which the rates and amounts of dropout differ by group, presents a significant threat to internal validity (Hansen, Collins, Malotte, Johnson, & Fielding, 1985). One specific concern is that those in the placebo

condition may tend to drop out more often than those in active treatment groups, perhaps due to a lack of treatment or side effects. If those in the placebo condition are more accurate than chance at judging their treatment assignment and consequently drop out of treatment, the extent and nature of blindness failure may be mischaracterized. Specifically, the remainder of participants in the placebo condition will be disproportionately comprised of incorrect judges, leading to a potentially spurious conclusion about blindness integrity. It is possible that those dropping out may have simply relapsed to smoking, and their judgment accuracy may not exceed chance or differ from those remaining in the study. Future studies should attempt to contact dropouts to determine the rates of judgment accuracy and the role of such judgments in the decision to drop out (Nordberg, 1992; Schulz & Grimes, 2002).

4.2.3. Additional strategies

Given the potentially significant impact of bias in clinical trials, a number of measures have been suggested to reduce the chances of the blind being broken (Even, Siobud-Dorocant, & Dardennes, 2000). Some have advocated for the use of "active" placebos that produce discernible sensations. Brownell and Stunkard (1982) showed that when tricyclic antidepressants are compared with either a placebo or an active drug (i.e., atropine) that produced a prominent side effect of tricyclics (i.e., dry mouth), fewer significant treatment effects were reported. Several nicotine replacement trials have used active placebos such as 0.5 or 1 mg nicotine gum (Arrechon & Punnontok, 1988; Clavel-Chapelon, Paoletti, & Benhamou, 1997; Hughes, Gust, Keenan, & Fenwick, 1990; Jarvis, Raw, Russell, & Feyerabend, 1982) and 1 or 3 mg in nicotine patch studies (Abelin et al., 1989; Richmond, Harris, & de Almeida Neto, 1994). However, it is unclear if such "active" placebos, especially the gum, have significant therapeutic effects if used enough (Leischow, Sachs, Hansen, & Bostrom, 1995). Another measure to enhance blindness is to compare multiple active treatments to placebo (Fisher & Greenberg, 1993) because participants are unlikely to be able to discriminate between active treatments (e.g., Margraf et al., 1991). Several NRT studies with three or more treatment conditions (e.g., Garvey et al., 2000; Herrera et al., 1995; Hughes et al., 1999; Jorenby et al., 1995; Tonnesen et al., 1988) should be meta-analyzed to determine the effects of multiple treatment groups on blindness integrity.

5. Conclusion

The NRT literature has been largely silent on the topic of blindness failure, even since the publication by Hughes and Krahn (1985) calling for researchers to address the problem. Based on the relatively few identified studies, definitive conclusions about the frequency and consequences of blindness failure are not justified. To determine the prevalence of failure, clinical trials of NRT should uniformly test the integrity of study blinds. Moreover, if blindness failure is observed, subsequent efforts should be made to determine if blindness failure is related to study outcome and, if so, to provide an estimate of treatment outcome adjusted for blindness bias. Without these methods and analyses, the validity of NRT clinical trial results could be questioned.

Acknowledgements

Earlier versions of this analysis were presented at the annual Minnesota Undergraduate Psychology Conference in 2001 and at the Society for Research on Nicotine Tobacco's annual conference in 2002. This article was supported by National Institute of Drug Abuse Grant P50-DA13333.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Abelin, T., Buehler, A., Muller, P., Vesanen, K., & Imhof, P. R. (1989). Controlled trial of transdermal nicotine patch in tobacco withdrawal. *Lancet*, *1*(8628), 7–10.

*Ahluwalia, J. S., McNagny, S. E., & Clark, W. S. (1998). Smoking cessation among inner-city African Americans using the nicotine transdermal patch. *Journal of General Internal Medicine*, *13*(1), 1–8.

*Anonymous (1993). Effectiveness of a nicotine patch in helping people stop smoking: Results of a randomised trial in general practice. Imperial Cancer Research Fund General Practice Research Group. *British Medical Journal*, *306*(6888), 1304–1308.

Arrechon, W., & Punnontok, J. (1988). Smoking cessation through the use of nicotine chewing gum: A double-blind trial in Thailand. *Clinical Therapeutics*, *10*, 183–186.

Brownell, K. D., & Stunkard, A. J. (1982). The double-blind in danger: Untoward consequences of informed consent. *American Journal of Psychiatry*, *139*, 1487–1489.

Buchkremer, G., Bents, H., Horstmann, M., Opitz, K., & Tolle, R. (1989). Combination of behavioral smoking cessation with transdermal nicotine substitution. *Addictive Behaviors*, *14*, 229–238.

Clavel-Chapelon, F., Paoletti, C., & Benhamou, S. (1997). Smoking cessation rates 4 years after treatment by nicotine gum and acupuncture. *Preventive Medicine*, *26*(1), 25–28.

Even, C., Siobud-Dorocant, E., & Dardennes, R. M. (2000). Critical approach to antidepressant trials. Blindness protection is necessary, feasible and measurable. *British Journal of Psychiatry*, *177*, 47–51.

Fisher, S., & Greenberg, R. P. (1993). How sound is the double-blind design for evaluating psychotropic drugs? *Journal of Nervous and Mental Diseases*, *181*, 345–350.

Friedman, L. M., Furburg, C. D., & DeMets, D. L. (1995). *Fundamentals of clinical trials*. New York: Springer.

Garvey, A. J., Kinnunen, T., Nordstrom, B. L., Utman, C. H., Doherty, K., Rosner, B., & Vokonas, P. S. (2000). Effects of nicotine gum by level of nicotine dependence. *Nicotine and Tobacco Research*, *2*, 53–64.

*Gourlay, S. G., Forbes, A., Marriner, T., Pethica, D., & McNeil, J. J. (1995). Double blind trial of repeated treatment with transdermal nicotine for relapsed smokers. *British Medical Journal*, *311*(7001), 363–366.

Greenberg, R. P., Bornstein, B. F., Greenberg, M. D., & Fisher, S. (1992). A meta-analysis of antidepressant outcome under "blinder" conditions. *Journal of Consulting and Clinical Psychology*, *60*, 664–669.

Hall, S. M., Delucchi, K. L., Velicer, W. F., Kahler, C. W., Ranger-Moore, J., Hedeker, D., Tsoh, J. Y., & Niaura, R. (2001). Statistical analysis of randomized trials in tobacco treatment: Longitudinal designs with dichotomous outcome. *Nicotine and Tobacco Research*, *3*(3), 193–202.

Hall, S. M., Tunstall, C. D., Ginsberg, D., Benowitz, N. L., & Jones, R. T. (1987). Nicotine gum and behavioral treatment: A placebo controlled trial. *Journal of Consulting and Clinical Psychology*, *55*(4), 603–605.

Hansen, W. B., Collins, L. M., Malotte, C. K., Johnson, C. A., & Fielding, J. E. (1985). Attrition in prevention research. *Journal of Behavioral Medicine*, *8*(3), 261–275.

Herrera, N., Franco, R., Herrera, L., Partidas, A., Rolando, R., & Fagerstrom, K. O. (1995). Nicotine gum, 2 and 4 mg, for nicotine dependence. A double-blind placebo-controlled trial within a behavior modification support program. *Chest*, *108*(2), 447–451.

Hjalmarson, A., Nilsson, F., Sjostrom, L., & Wiklund, O. (1997). The nicotine inhaler in smoking cessation. *Archives of Internal Medicine*, *157*(15), 1721–1728.

Hughes, J. R., Gust, S. W., Keenan, R. M., & Fenwick, J. W. (1990). Effect of dose on nicotine's reinforcing, withdrawal-suppression and self-reported effects. *Journal of Pharmacology and Experimental Therapeutics*, *252*(3), 1175–1183.

Hughes, J. R., Gust, S. W., Keenan, R. M., Fenwick, J. W., & Healey, M. L. (1989). Nicotine vs. placebo gum in general medical practice. *Journal of the American Medical Association*, *261*(9), 1300–1305.

Hughes, J. R., Gust, S.W, & Pechacek, T. F. (1987). Prevalence of tobacco dependence and withdrawal. *American Journal of Psychiatry*, *144*, 205–208.

Hughes, J. R., Hatsukami, D. K., Pickens, R. W., Krahn, D., Malin, S., & Luknic, A. (1984). Effect of nicotine on the tobacco withdrawal syndrome. *Psychopharmacology*, *83*(1), 82–87.

Hughes, J. R., & Krahn, D. (1985). Blindness and the validity of the double-blind procedure. *Journal of Clinical Psychopharmacology*, *5*(3), 138–142.

Hughes, J. R., Lesmes, G. R., Hatsukami, D. K., Richmond, R. L., Lichtenstein, E., Jorenby, D. E., Broughton, J. O., Fortmann, S. P., Leischow, S. J., McKenna, J. P., Rennard, S. I., Wadland, W. C., & Heatley, S. A. (1999). Are higher doses of nicotine replacement more effective for smoking cessation? *Nicotine and Tobacco Research*, *1*, 169–174.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Jarvis, M. J., Raw, M., Russell, M. A., & Feyerabend, C. (1982). Randomised controlled trial of nicotine chewing-gum. *British Medical Journal Clinical Research Ed.*, *285*(6341), 537–540.

Jorenby, D. E., Smith, S. S., Fiore, M. C., Hurt, R. D., Offord, K. P., Croghan, I. T., Hays, J. T., Lewis, S. F., & Baker, T. B. (1995). Varying nicotine patch dose and type of smoking cessation counseling. *Journal of the American Medical Association*, *274*(17), 1347–1352.

Leischow, S. J., Sachs, D. P., Hansen, M. D., & Bostrom, A. G. (1995). Nicotine polacrilex dose effects: Serum nicotine levels and sensory characteristics. *Outcomes Management*, *117*(2), 125–129.

*Lin, H. (1996). The effectiveness of nicotine patch for smoking cessation. *Chinese Psychiatry*, *10*, 223–233.

*Killen, J. D., Fortmann, S. P., Davis, L., & Varady, A. (1997). Nicotine patch and self-help video for cigarette smoking cessation. *Journal of Consulting and Clinical Psychology*, *65*(4), 663–672.

Margraf, J., Ehlers, A., Roth, W. T., Clark, D. B., Sheikh, J., Agras, W. S., & Tayor, C. B. (1991). How "blind" are double-blind studies? *Journal of Consulting and Clinical Psychology*, *59*, 184–187.

Microsoft Access. (1997). Access 97 (Access 97). Redmond, WA: Microsoft Corp.

Microsoft Excel. (1997). Excel 97 (Excel 97). Redmond, WA: Microsoft Corp.

Nash, H. (1957). The double-blind procedure: Rationale and empirical evaluation. *Journal of Nervous and Mental Disease*, *134*(1), 34–47.

Nordberg, P. M. (1992). Leave no stone unturned. *American Journal of Epidemiology*, *136*(9), 1160–1166.

Orwin, R. G. (1994). Evaluating coding decisions. In H. Cooper, & L. Hedges (Eds.), *The handbook of research synthesis* (pp. 139–162). New York: Russel Sage Foundation.

Oxtoby, A., Jones, A., & Robinson, M. (1989). Is your 'double-blind' design truly double-blind? *British Journal of Psychiatry*, *155*, 700–701.

Perkins, K. A., DiMarco, A., Grobe, J. E., Scierka, A., & Stiller, R. L. (1994). Nicotine discrimination in male and female smokers. *Psychopharmacology (Berl)*, *116*(4), 407–413.

Perkins, K. A., D'Amico, D., Sanders, M., Grobe, J. E., Wilson, A., & Stiller, R. L. (1996). Influence of training dose on nicotine discrimination in humans. *Psychopharmacology*, *126*(2), 132–139.

Perkins, K. A., Sanders, M., D'Amico, D., & Wilson, A. (1997). Nicotine discrimination and self-administration in humans as a function of smoking status. *Psychopharmacology (Berl)*, *131*(4), 361–370.

Richmond, R. L., Harris, K., & de Almeida Neto, A. (1994). The transdermal nicotine patch: Results of a randomised placebo-controlled trial. *Medical Journal of Australia*, *161*(2), 130–135.

SAS. (2002). *Statistical analysis system (Version 8.01)*. Cary, NC: SAS Institute.

Schneider, N. G., Olmstead, R., Mody, F. V., Doan, K., Franzon, M., Jarvik, M. E., & Steinberg, C. (1995). Efficacy of a nicotine nasal spray in smoking cessation: A placebo-controlled, double-blind trial. *Addiction*, *90*(12), 1671–1682.

Schulz, K. F., & Grimes, D. A. (2002). Sample size slippages in randomised trials: Exclusions and the lost and wayward. *Lancet*, *359*(9308), 781–785.

*Sonderskov, J., Olsen, J., Sabroe, S., Meillier, L., & Overvad, K. (1997). Nicotine patches in smoking cessation: A randomized trial among over-the-counter customers in Denmark. *American Journal of Epidemiology*, *145*(4), 309–318.

Tonnesen, P., Fryd, V., Hansen, M., Helsted, J., Gunnersen, A. B., Forchammer, H., & Stockner, M. (1988). Effect of nicotine chewing gum in combination with group counseling on the cessation of smoking. *New England Journal of Medicine*, *318*(1), 15–18.

Tonnesen, P., Norregaard, J., Mikkelsen, K., Jorgensen, S., & Nilsson, F. (1993). A double-blind trial of a nicotine inhaler for smoking cessation. *Journal of the American Medical Association*, *269*(10), 1268–1271.

*Tonnesen, P., Norregaard, J., Simonsen, K., & Sawe, U. (1991). A double-blind trial of a 16-hour transdermal nicotine patch in smoking cessation. *New England Journal of Medicine*, *325*(5), 311–315.

*Westman, E. C., Levin, E. D., & Rose, J. E. (1993). The nicotine patch in smoking cessation. A randomized trial with telephone counseling. *Archives of Internal Medicine*, *153*(16), 1917–1923.