Nicotine & Tobacco Research (2003) **5,** 13–25

Taylor & Francis
healthsciences

Review

Measures of abstinence in clinical trials: issues and recommendations

John R. Hughes, Josue P. Keely, Ray S. Niaura, Deborah J. Ossip-Klein, Robyn L. Richmond, Gary E. Swan

A workgroup formed by the Society for Research on Nicotine and Tobacco reviewed the literature on abstinence measures used in trials of smoking cessation interventions. We recommend that trials report multiple measures of abstinence. However, at a minimum we recommend that trial: (a) report prolonged abstinence (i.e., sustained abstinence after an initial period in which smoking is not counted as a failure) as the preferred measure, plus point prevalence as a secondary measure; (b) use 7 consecutive days of smoking or smoking on $\geq$1 day of 2 consecutive weeks to define treatment failure; (c) include non-cigarette tobacco use, but not nicotine medications in definitions of failure; and (d) report results from survival analysis to describe outcomes more fully. Trials of smokers willing to set a quit date should tie all follow-ups to the quit date and report 6- and/or 12-month abstinence rates. For these trials, we recommend an initial 2-week grace period for prolonged abstinence definitions; however, the period may vary, depending on the presumed mechanism of the treatment. Trials of smokers who may not be currently trying to quit should tie follow-up to the initiation of the intervention and should report a prolonged abstinence measure of $\geq$6-month duration and point prevalence rates at 6- and 12-month follow-ups. The grace period for these trials will depend on the time necessary for treatment dissemination, which will vary depending on the treatment, setting, and population. Trials that use short-term follow-ups ($\leq$3 months) to demonstrate possible efficacy should report a prolonged abstinence measure of $\geq$4 weeks. We again recommend a 2-week grace period; however, that period can vary.

Introduction

The Society for Research on Nicotine and Tobacco (SRNT) formed a workgroup, including a subcommittee on abstinence outcome measures, to examine outcome measures used in clinical trials. The subcommittee defined its charge as examining: (a) continuous abstinence, prolonged abstinence, sustained abstinence, point prevalence and repeated point-prevalence measures and the use of grace periods; (b) definitions of treatment failure; (c) whether non-cigarette tobacco use and non-tobacco nicotine use should be termed failures; (d) short-term (1–3 months), vs. long-term (6 and 12 months), vs. very-long-term (>12 months) follow-ups; and (e) non-traditional measures (e.g., survival analysis-based measures). Other workgroups reported on biochemical verification (Benowitz et al., 2002), withdrawal (Shiffman, West, & Gilbert, 2000), harm reduction (Henningfield et al., 2000), adolescent outcome measures (Mermelstein et al., 2002) and statistical issues (Hall et al., 2001).

We gathered information via a literature search of *Medline* and *Psych Abstracts* and the Office on Smoking and Health's *Bibliography on Smoking and Health*, collections of the authors, references in review and methodological articles, and requests on the SRNT e-mail list. This search produced several reviews and meta-analyses that commented on methodological issues (Cohen et al., 1989; Grabowski & Bell, 1983; Lando; 1983; Midanik, Polen, Hunkeler, Tekawa, &

John R. Hughes (chair), Josue P. Keely, Department of Psychiatry, Psychology and Family Practice University of Vermont, USA; Ray S. Niaura, Division of Behavioral Medicine, Miriam Hospital, Providence RI, USA; Deborah J. Ossip-Klein, Smoking Research Program, University of Rochester Cancer Center, NY, USA*; Robyn L. Richmond, School of Community Medicine, University of New South Wales, Australia*; Gary E. Swan, Center for Health Sciences, SRI International Mealo Park, CA, USA*.

Correspondence: John R. Hughes, MD, University of Vermont, Department of Psychiatry*, 38 Fletcher Place, Burlington VT 05401-1419, USA. Tel: +1 (802) 656-9610; fax: +1 (802) 656-9628; e-mail: john.hughes@uvm.edu

ISSN 1462-2203 print/ISSN: 1469-994X online/02/02013504
DOI: 10.1080/1462220031000070552

Soghikian, 1985; Miller, 1996; Ossip-Klein, Bigelow, Parker, Curry, Hall, & Kirkland, 1986; Richmond, 1997; Shipley, Rosen, & Williams, 1982; Stapleton, 1998; Velicer, Prochaska, Rossi & Snow, 1992). The reviews by Velicer, Prochaska, Rossi & Snow (1992), Miller (1996), Ockene et al., 2000, and Stapleton (1998)) are especially relevant and are recommended reading. Many other clinical trials have made methodological comments on outcome measures, but may have been missed in our review because they did not include methodological terms in their titles or keywords.

Although unwritten conventions exist about reporting abstinence in clinical trials, no recent written guideline is well accepted. In the 1970s and 1980s, the American Lung Association, the International Union Against Cancer, the U.S. National Interagency Council on Smoking or Health, and the US National Center for Health Education each issued guidelines (Lando, 1983), but these guidelines have been cited only rarely. In 1986, the U.S. National Heart Lung and Blood Institute proposed a definition of relapse that has been applied to many clinical trials (Ossip-Klein et al., 1986). In the 1990s, the U.S. Food and Drug Administration (FDA) began using a single definition of abstinence (described below) in registration trials of smoking cessation pharmacotherapies (U.S. FDA, 1995). The World Health Organization defined former smokers as those who have not smoked in the last year (Ramstrom, 1987); this definition can be interpreted to mean that minimum of 1 year should be used to define abstinence. In 1999, the College on Problems of Drug Dependence (CPPD) and the National Institute on Drug Abuse (NIDA) together suggested outcome measures for clinical trials in drug dependence, including nicotine dependence (CPPD, 1999). In 2000, Ockene et al. also recommended definitions of *lapse*, *relapse* and *short-* and *long-term cessation*. The current article adds to these guidelines by discussing the relative pros and cons of the different abstinence measures in detail and by examining how abstinence measures may differ across types of interventions or populations being studied.

Definition of terms

This section defines several terms used in the paper. *Abstinence* without a qualifier refers to a period in which there is no relapse. *Relapse* refers to a return to regular smoking after some period of abstinence. *Grace period* is a period immediately after the quit date or intervention in which continued smoking is not counted as a failure. *Outcome* refers only to the dependent measure in the trial, not efficacy. *Efficacy*, *treatment effect* and *effect size* refer to the difference in outcomes between the experimental and the control group. *Failure* refers to an outcome other than the goal of the treatment. *Clinical trial* refers to any prospective study of an intervention to increase abstinence (i.e., it is not restricted to controlled trials, to trials of smokers currently trying to quit, or to trials of pharmacotherapy). These and other terms are defined in more detail when they are discussed below.

Typical use of abstinence measures in clinical trials

Many reasons exist for determining abstinence, e.g., to assess the relationship of smoking and health outcomes, to track the effects of tobacco control programs and to determine whether a clinical intervention is effective. The definition of abstinence varies to match the reason for its use. For example, epidemiological studies often require 1–2 years of abstinence to avoid falsely declaring a temporary abstainer as a former smoker (Midanik et al., 1985), whereas an initial trial of a new clinical intervention may require only 4 weeks of abstinence. Our focus here is solely on the use of abstinence in trials of clinical interventions either to prompt or to aid in achieving life-long abstinence.

For clarity, we have divided such clinical trials into three types. *Aid-to-cessation* trials test a treatment among smokers currently willing to quit; group behavior therapy and medication trials are examples. In most of these trials, the experimenter, not the subject, determines a quit date and delivers the treatment before and/or shortly after the quit date. Unlike most treatment studies, which tie follow-up to the end of treatment, smoking studies often tie follow-up times to a quit date (Figure 1). Thus, it is

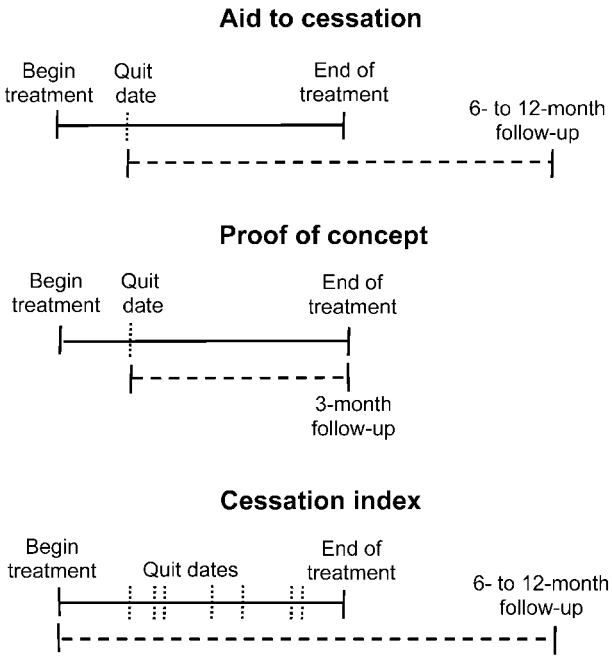

Figure 1. Relation of follow-up to quit dates and end-of-treatment across different types of clinical trials.

important to define the quit date clearly, either prospectively or retrospectively, especially when multiple self-selected quit dates need to be accommodated.

Most, but not all, subjects quit at least for a day on the quit day (a *quit*) (Hughes et al., 1992). After this, the large majority back to smoking. Often smoking begins with a few cigarettes (a *lapse*) and rapidly progresses to regular smoking (a *relapse*). The curve describing the number of subjects remaining abstinent over time (a survival or relapse curve) after a given quit attempt finds the number of continuing abstainers decreases over time (Figure 2). Usually the treatment group does better early on and maintains that advantage over time; however, sometimes the initial treatment effect dissipates over time. Once a subject either lapses or relapses, he/she often is dropped from the study as a treatment failure.

Proof-of-concept trials are similar, but test only whether of not a treatment is likely to achieve efficacy. These trials use a short follow-up ($\leq$3 months). Most clinicians, scientists and administrators would require longer-term follow-up to be convinced of efficacy.

Cessation-induction trials test a treatment to prompt cessation among all smokers, including those who are not currently trying to quit. Physician advice for all patients seen and telephone counseling for all smokers enrolled in an HMO are examples. Subjects usually come into contact with the treatment at irregular times, which may be weeks or months after the intervention is begun. In addition, subjects may contact the intervention at several points during the study. The intervention often is designed to increase the probability of quit attempts during some period of time, with many attempts resulting in abstinence. As a result, a successful intervention typically increases the number of abstainers over time as it motivates more and more smokers to quit, with some continuing abstinence (Fig. 1). To accommodate these types of outcomes, these trials typically focus on percent abstinent at a given time and do not tie a follow-up to a determined quit date. Instead, they tie follow-up to date of onset or of offset of their intervention and use abstinence measures that count later quit episodes as successes.

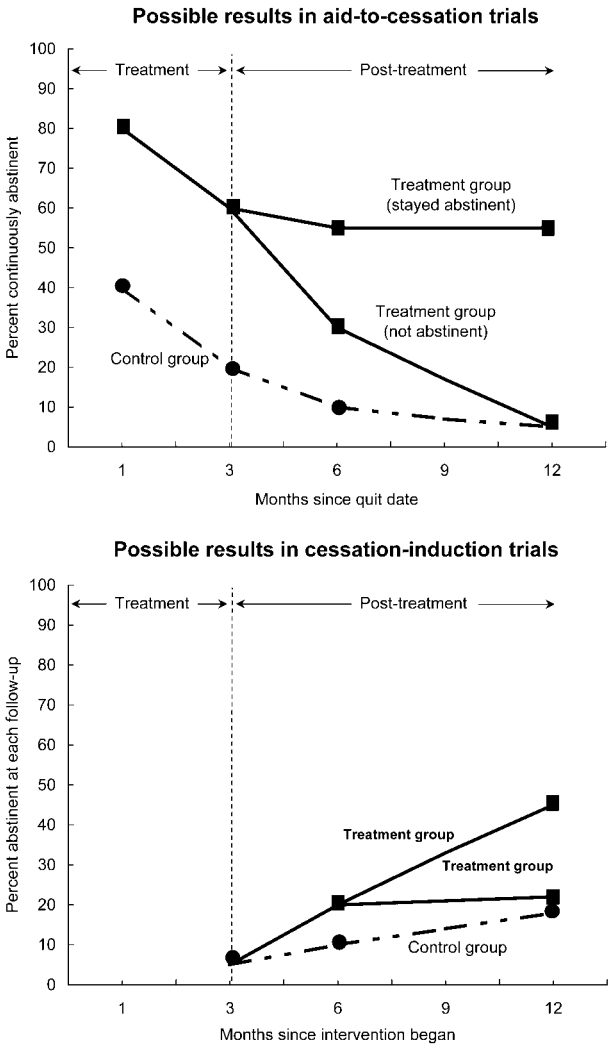

Figure 2. Outcome curves showing possible outcomes for two types of studies.

Evaluating and interpreting measures of abstinence

It is difficult to know what criteria to use in judging whether or not a certain abstinence measure is better than another measure. Smoking cessation trials are usually conducted to induce life-long abstinence from cigarettes to decrease the risks of health disorders [U.S. Department of Health and Human Services [USDHHS], 1990. Thus, one could argue that the ability of the measure to be a proxy for life-long abstinence should be a major judgment criterion. Others would argue that in the context of a clinical trial, a major goal of the abstinence measure is to quantify the treatment effect size accurately, i.e., the magnitude of the difference between active and control conditions via an odds ratio, or the absolute difference in percent abstinence at the primary follow-up.

Abstinence outcomes themselves (e.g., percentage abstinent) often are used as a direct measure of treatment efficacy. We believe that doing so is problematic because meta-analyses suggest that absolute abstinence rates can vary widely, depending on several non-treatment factors such as population, setting, amount of adjunctive therapy, length of follow-up and definitions of abstinence (Covey & Glassman, 1991; Fiore et al., 1994; Fisher, Glasgow, & Terborg, 1990; Gourlay & Benowitz, 1995; Gourlay & McNeil, 1990; Law & Tang, 1995; Po, 1993; Silagy Mant, Fowler & Lodge, 1994a, b; Tang, Law, & Wald, 1994). These same meta-analyses suggest that efficacy or the effect size (i.e., the difference in or ratio of abstinence in the experimental vs. in the control group) is influenced much less by these variables. This

would argue that some measure of effect size (not the abstinence measure) should be the major criterion for success.

A word of caution. We found that empirical data on the pros and cons of different abstinence measures were available in only a handful of papers; thus, logic, clinical wisdom and consensus were major factors in our decisions. *As a result, readers should focus as much on understanding our arguments provided as on our final recommendations.*

Continuous, prolonged, sustained, point prevalence, and repeated point prevalence abstinence and the use of grace periods

Velicer and colleagues (1992) have provided an excellent discussion of continuous abstinence, prolonged abstinence, point prevalence and repeated point prevalence measures, and many of the arguments below are drawn from that discussion; in addition, this group has presented empirical data on the relationship among these measures (Velicer et al., in press). Prior papers have often used these terms interchangeably, which has led to confusion. Figure 3 illustrates our definitions of these terms. *Continuous abstinence* refers to abstinence between quit day and a follow-up time; *prolonged abstinence* refers to sustained abstinence after an initial grace period or to a period of sustained abstinence between two follow-ups (the two are equivalent); *point prevalence abstinence* refers to the prevalence of abstinence during a time window immediately preceeding follow-up; and *repeated point prevalence* refers to point prevalence abstinence at two or more follow-ups between which smoking is allowed. These terms, especially *continuous abstinence*, have often been assumed to refer to absolutely no smoking. In contrast, as discussed below, our use of these terms assumes only that a definition of relapse or failure is not met. Because these definitions may not be ones readers use themselves, we urge readers to remember these definitions as they read our text. In all these definitions, the usual practice is to treat those lost to follow-up as smokers. The pros and cons of this practice are discussed in the SRNT methods paper by Hall et al. (2001).

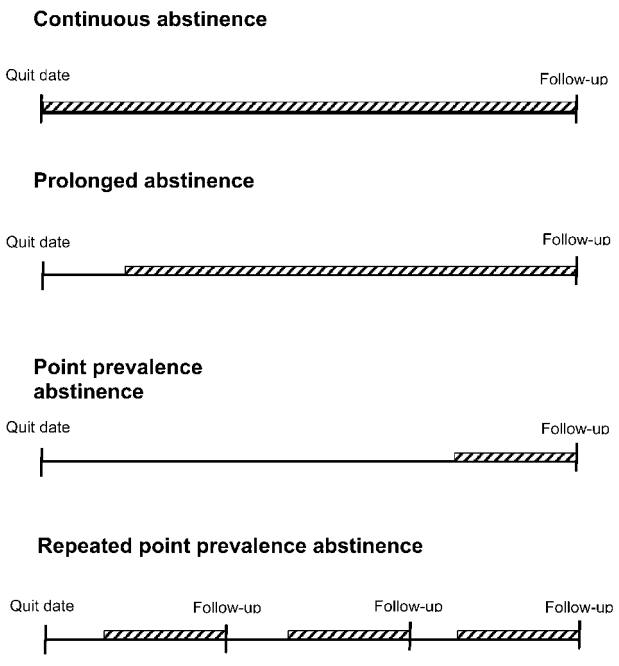

Figure 3. Relation of duration of abstinence to follow-up time across different abstinence measures.

Early meta-analyses noted that more studies reported point prevalence than continuous abstinence or prolonged abstinence measures (Fiore et al., 1994), but more recent meta-analyses have suggested the opposite (Fiore et al., 2000). Repeated point prevalence is probably the least commonly used measure. Different measures can and often do produce different rates of abstinence. By definition, at any given follow-up, point prevalence $\geq$ repeated point prevalence $\geq$ prolonged abstinence $\geq$ continuous abstinence.

In addition, by definition, point prevalence can increase over time, but continuous abstinence, prolonged abstinence and repeated point prevalence cannot. Several reviews have compared the absolute rates using these measures across several studies (Velicer, 1992). For example, a collection of 10 studies of self-quitting reported a mean point prevalence abstinence at 6 months after the quit date of 13%, whereas 6-month rates for continuous abstinence and repeated point prevalence using the same criteria for amount of smoking were 5%–6%. The pattern for 12-month follow-up was similar (Cohen et al., 1989).

Studies also have examined whether the effect size (i.e., experimental vs. control comparison) varies across these measures (Table 1). Although these results suggest that continuous abstinence or prolonged abstinence measures produce larger treatment effects than do point prevalence measures, the number of meta-analytic tests of continuous abstinence/prolonged abstinence vs. point prevalence is small.

Continuous abstinence and prolonged abstinence. Many have considered continuous abstinence to be the gold standard — probably because continuous abstinence requires a longer period of abstinence than the other measures and thus is more likely to represent long-term abstinence. In addition, because it is tied to the quit date, continuous abstinence is more logically tied to intent-to-treat conventions. Finally, because continuous abstinence is the most rigorous and most conservative measure, it has been thought to be the fairest.

Table 1. Meta-analyses that reported effect sizes by abstinence measure

Meta-analysis (follow-up)	Treatment	Measure	Abstinence measure		
			PP	*PA*	*CA*
Fiore et al., 1994 (3-month)	Nicotine patch	*OR*	2.6	–	3.8
Fiore et al., 1994 (6-month)	Nicotine patch	*OR*	2.6	–	3.2
Fisher et al., 1990 (12-month)	Worksite treatments	*d*	"Higher effect size using CA" (vs. PP)[b]		
Richmond, 1997 (6- or 12-month)	Nicotine patch	Fraction of positive studies[a]	2/6	4/4	3/3

PP, point prevalence; *PA*, prolonged abstinence; *CA*, continuous abstinence; *OR*, odds ratio; *d*, probit change score.
[a]Statistically significant advantage of nicotine patch.
[b]$p < .05$.

One argument against continuous abstinence and for prolonged abstinence has been the lack of a grace period with continuous abstinence. Although most who slip during initial abstinence return to regular smoking (Kenford, Fiore, Jorenby, Smith, Wetter, & Baker, 1994), a substantial minority of smokers who eventually quit for their lifetime smoke a few cigarettes in the first days of abstinence (Hughes et al., 1992). For example, in one study of self-quitters, 23% of those who quit continuously between 3 and 6 months had a few cigarettes early on (Hughes et al., 1992). Thus, one rationale for allowing grace periods (i.e., for prolonged abstinence) is that they will include persons who will achieve life-long abstinence, but who would have been counted as failures by a no-grace period continuous abstinence measure. Whether or not use of a grace period influences the effect size in a clinical trial has not been tested.

Another rationale for grace periods is that although existing treatments appear to work very soon after initiation, it is possible that some treatments might have delayed effects. For example, a main effect of naltrexone for alcoholism is that it decreases the probability of slip converting to regular drinking (Anton, Moak, Waid, Latham, Malcolm, & Dias, 1999). An early analysis suggests bupropion might similarly serve as a rescue therapy (Jamerson et al., 2001). If one allowed an initial grace period during which slips could be rescued and counted as successes, one could detect this effect. If one did not allow a grace period, one could not detect this effect.

In fact, whether any treatment is fully effective beginning on the quit day can be questioned. For example, nicotine replacement therapy (NRT) is typically begun only on the first day of abstinence. Given the need for cumulative dosing and the learning that must occur with participant-determined dosing, it is unlikely that a full pharmacological effect of NRT would occur by the end of the quit day. Similar arguments could be made for behavioral therapy, given the prequit date. Thus, some grace period, even if it is short, could be argued to be appropriate for all treatments.

The duration for a grace period that one should use is unclear. Clinical intuition is that smokers can more easily recover from early slips (e.g., in the first week after quitting) and reinitiate abstinence than they can recover from later slips (e.g., during the third to fourth week after quitting); however, we are unaware of any empirical tests of this notion. Currently, in deciding whether or not to approve smoking cessation medications, the FDA uses a grace period with no smoking at all for the next 4 weeks.

Most grace periods focus on days of smoking after quitting. An alternative is to include the amount smoked (e.g., smoking of less than a certain amount on any given day or less than some total number of cigarettes). Although clinical intuition is that the number of days of smoking is more predictive of failure than the amount smoked on any one day, this assumption has not been empirically tested. Finally, one argument against both continuous abstinence and prolonged abstinence is that many smokers' memories about smoking before a few weeks ago may be poor.

Point prevalence and repeated point prevalence abstinence measures. One argument for point prevalence measures is similar to that for inclusion of grace periods; that is, they can capture delayed effects of an intervention. Such sleeper effects have not been documented very often in aid-to-cessation trials (Ockene et al., 1994), but they have been clearly shown in cessation-induction studies with expert/tailored interventions (Prochaska, DiClemente, Velicer, & Rossi, 1993), receipt of a prescription for pharmacotherapy (Russell, Merriman, Stapleton, & Taylor, 1983), and other treatments.

Another argument for point prevalence is that continuous abstinence or prolonged abstinence measures assume that once someone fails he/she is unlikely to return to long-term abstinence during the remainder of the study (i.e., little recycling occurs). The only study of which we are aware that actually followed subjects after a relapse found that this assumption may not be correct. Of those who relapsed after

some period of abstinence, 43% went on to achieve prolonged abstinence soon thereafter (Jarvis, West, Tunstall-Pedoe, & Vesley, 1984).

A last argument for point prevalence is that a recent psychometric analysis of continuous abstinence, prolonged abstinence, and point prevalence measures concluded that a point prevalence measure may be the preferred outcome measure: When these outcomes were compared, point prevalence was the measure most highly correlated with other measures; that is, it had the highest concurrent validity (Velicer & Prochaska, 2001).

The major argument against point prevalence measures is that because the duration of abstinence required to fulfill the measure is small, this measure should be a poorer predictor of long-term abstinence. In addition, point prevalence provides a more heterogeneous sample, including smokers who have been abstinent for only a week and those who have been abstinent for many months. As a result, point prevalence should be less stable than continuous abstinence. This would suggest continuous abstinence or prolonged abstinence is a better measure for meta-analyses of treatment efficacy than point prevalence, given the recurring problem of across-study and across-follow-up variability in meta-analyses.

Point prevalence has used several time windows, with 7 days being the most common. One major issue with choosing a window is choosing one that is biochemically verifiable. The SRNT Biochemical Verification Workgroup has stated that the biochemically verifiable windows are 7 days for cotinine, 1 day for carbon monoxide and 28 days for thiocyanate (Benowitz et al., 2002).

Repeated point prevalence is used to overcome some of the problems with point prevalence mentioned above. One rationale behind repeated point prevalence is that it would be unusual to find a smoker who was abstinent at three consecutive follow-ups, but smoked in between. In addition, repeated point prevalence is an estimate of abstinence over time that includes biochemically verifiable self-reports.

Recommendations

In terms of definitions, we recommend that, to prevent confusion with prolonged abstinence, the term *continuous abstinence* should refer only to abstinence periods that begin on the quit date. We also recommend not using the term *sustained abstinence* because its definition is unclear and is redundant with other terms.

In terms of methodology, we recommend that follow-ups for aid-to-cessation and proof-of-concept trials be tied to the quit date, because this approach has been the tradition and is convenient for survival analysis (see below). We suggest that follow-ups for cessation-induction studies should be tied to the onset of intervention, because it is problematic to tie them to multiple quit dates. We recommend onset rather that offset of intervention, because this more closely parallels the timing of follow-ups in aid-to-cessation trials.

We recommend prolonged abstinence as the preferred measure because: (a) it requires a long period of abstinence, (b) it captures long-term abstainers who initially slip and (c) it can be used with treatments that have a delayed effect. In addition, although grace periods were initially developed for aid-to-cessation and proof-of-concept trials, the concept can be applied to cessation-induction trials to cover the period during which the intervention is being disseminated and during which one would not expect to see large treatment effects. This period can be several months in duration.

For most aid-to-cessation and proof-of-concept studies, we recommend a 2-week grace period based on days of smoking, not the amount smoked each day. This grace period, which is based solely on tradition and clinical observation, may need to vary depending on the presumed mechanism of action of the treatment. For cessation-induction studies, we cannot recommend a precisely defined grace period because the time it takes for treatment dissemination can vary widely depending on population, setting, treatment, etc.

Many have considered continuous abstinence, not prolonged abstinence, as the gold standard, and we are not necessarily recommending against continuous abstinence. If an investigator has a good, a priori reason to believe the treatment will be essentially fully effective on the quit day, we recommend he/she use a prolonged abstinence with no grace period, i.e., continuous abstinence.

We recommend that aid-to-cessation and cessation-induction trials also report point prevalence with a 7-day window as a secondary measure because: (a) many prior trials have reported only point prevalence, and many meta-analyses have been and will be based on point prevalence; and (b) the 7-day window is the most common and can be verified by blood or salivary cotinine. If a trial uses a non-cotinine method of verification, we suggest it also report point prevalence with a verifiable window. We also recommend reporting point prevalence with a 30-day window to allow comparison with trials in adolescent smokers (see SRNT Adolescent Workgroup, Mermelstein et al., 2002). For proof-of-concept trials, we see no need for reporting point prevalence because cross-study analyses typically do not include these short-term trials. Finally, we recommend not using the term repeated point prevalence because we believe it adds little over prolonged abstinence.

Definition of failure *for prolonged abstinence and continuous abstinence measures*

Because the first smoking cessation trials were aid-to-cessation trials in which most smokers quit at least initially, participants who fail (also called *non-responders*, *protocol violators*, etc.) were conceptualized as *relapsers*. This has caused some confusion because definitions of the term *relapse* require some period of initial abstinence. However, some smokers never quit in an aid-to-cessation trial. Such smokers are clearly failures yet are technically not relapsers because they have never quit. Thus, for the purposes of this review we will use the term *failure* instead of *relapse*.

This is not to say that relapse itself should not be investigated. For example, some behavioral therapy and medication treatment trials have tested the specific hypothesis that a treatment will not improve initial abstinence but will specifically decrease later relapse (Killen, Fortmann, Newman, & Varady, 1990). Thus, in some cases, researchers may wish to conduct a separate analysis specifically about the effect of the treatment on relapse. In that case, a clear definition of an initial quit is needed. The most commonly used one is the NHLBI relapse definition for which a *quit* is defined as self-reported no smoking at all for 24 hr.

How much use (in terms of frequency and duration) is necessary for a failure? The two most commonly used thresholds have been the not-even-a-puff criterion and the NHLBI definition of relapse (i.e., any smoking on 7 consecutive days) (Ossip-Klein et al., 1986). Many other definitions of failure (e.g., the NHLBI lapse criterion, smoking an average of one cigarette per day since the last follow-up and smoking for 4 days in the last week) have been used. We are unaware of tests of which definitions best predict the inability to regain life-long abstinence. Studies that have used both a not-even-a-puff and a slips-allowed criterion found similar effect sizes for both measures (Schneider et al., 1995; Sutherland et al., 1992; Tonnesen, Norregaard, Simonsen, & Sawe, 1991).

One major asset of the not-even-a-puff criterion (or the NHLBI lapse criterion) is that, depending on the duration of abstinence claimed, it is possible to verify this criterion biochemically. One liability of this criterion is that it assumes that 100% of those who take a puff will go on to regular smoking. As described earlier, this is not the case. However, we currently do not know the shape of the quantitative relationship between the number of cigarettes smoked or the number of days smoked and the probability of eventual failure/success. Thus, any cutoff used is based on clinical judgment. Finally, most clinicians are more interested in a return to prolonged daily smoking than a return to smoking just one puff.

The major asset of the NHLBI definition of relapse as a criterion for failure is that it requires repeated use of tobacco (i.e., does not count isolated slips as failure). One problem with the NHLBI definition of relapse is that it is not biochemically verifiable; for example, if a subject stated that he/she had not smoked except on the day before and the day of the follow-ups, he/she would not meet the NHLBI criterion for failure and thus would be considered a success but would produce a positive biochemical value. Another potential problem with the NHLBI definition is that, at the time the definition was made, little information was available about non-daily smokers, and researchers assumed that such smoking was very rare. However, recent surveys have shown that many abstinent smokers become non-daily smokers (Borland, 1994; Gerlach et al., 1998; Gilpin, Cavin & Pierce, 1997). Such chronic occasional smokers could be defined as abstinence successes by the NHLBI criterion.

Definitions of failure for point prevalence have used retrospective windows of various durations, but 7 days has been the most common window employed. The NHLBI definition could be discrepant with a point prevalence measure with a 7-day window, because with the NHLBI definition, smoking on 6 of the 7 days of a window could still count as a success.

Recommendations

We recommend the definition of failure for prolonged abstinence and continuous abstinence in clinical trials should require smoking on several occasions; thus, we recommend the definition of failure should fulfill the NHLBI definition (smoking on 7 consecutive days) or smoking at least once each week over 2 consecutive weeks. The latter part of this definition is included to count those who smoke on a regular but less than daily basis as a failure. We are not recommending the not-even-a-puff criterion because it is overly stringent in that it counts as failures long-term abstainers who have a single slip. We recognize that our definition of failure is not always biochemically verifiable; however, in reality, verifying any definition of long-term abstinence would involve conducting multiple biochemical tests (e.g., a cotinine level every 2 weeks for 6 months), which thus far has not been accomplished.

For point prevalence, we recommend that the definition of failure for this measure should be any smoking (even a puff) during a 7-day window. We acknowledge that, in terms of the 7 days before follow-up, the criterion for point prevalence is actually more stringent than for prolonged abstinence and continuous abstinence measures. However, because point prevalence is intended to be a cross-sectional snapshot, using this more stringent window is appropriate.

Non-cigarette tobacco and non-tobacco nicotine use during cigarette abstinence

Most studies of tobacco treatments have focused on cigarettes because they are the most prevalent and most deadly form of tobacco (USDHHS, 1990). However, in some countries (e.g., Sweden), some smokers begin using non-cigarette tobacco (Jarvis et al., 1984). In the U.S., this is rare (Gerlach et al., 1998).

One choice for dealing with non-cigarette tobacco is to ignore it. The rationale here is that the treatment claims only the induction of abstinence from cigarettes, plus the health risks of most non-cigarette tobacco use appear to be much less than those of cigarette use (Shanks & Burns, 1998). On the other hand, abstinent smokers inhale smoke from cigars and pipes deeply, and are at increased risk compared with pipe/cigar smokers who have never used cigarettes (Shanks & Burns, 1998).

The other choice is to define use of any type of tobacco as a failure. The rationale here is that most smokers stop smoking to improve their health; thus, a study should define success as eliminating all tobacco risks. Since much cigar and pipe use involves, non-daily chronic use (Gerlach et al., 1998), it would be important to include this tobacco use in the failure defined as any use in 2 consecutive weeks.

Another issue is the use of non-tobacco nicotine (i.e., how to handle continuing use of NRT during abstinence). One method is to ignore NRT use. The rationale here is that most smokers stop smoking to eliminate tobacco-related risks, and long-term use of NRT has not been found to be harmful (Benowitz, 1998).

A second method is to consider NRT users to be failures. The rationale is that if the treatment claims to treat nicotine dependence and if a subject continues to use nicotine in some form, one cannot claim that subject as a success. This debate is similar to the debate concerning methadone. One significant difference between long-term NRT use and methadone use is that most long-term users of NRT do not appear to be physically or psychologically dependent on nicotine, but rather simply believe that they need a longer period of treatment (Hughes, 1998). A third solution is offered by the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders: Fourth Edition* (1994) nosology, which distinguishes between those abstinent and on agonist therapy, and those abstinent and not on agonist therapy.

Recommendations

Because the harms of non-cigarette tobacco are significant and because many ex-smokers inhale smoke from cigars and pipes, we recommend that the definition of failure should include non-cigarette tobacco use. In contrast, because the risks of long-term NRT appear small, we recommend that definitions of failure should not include non-tobacco nicotine use.

Length of follow-up

One criterion to use in determining the necessary length of follow-up is how closely the actual percentage abstinent corresponds with life-long abstinence. Logically, the longer the follow-up, the closer it should correspond to the life-long pattern. The question is at what point longer follow-ups are worthwhile. Enough relapse curves have been published to make it clear that substantial relapse occurs through the first year after the quit date. Most studies have suggested that non-trivial relapse occurs even after 1 year; others have not (Becona & Vazquez, 1998; Brandon, Lazev, & Juliano, 1998; Richmond, 1997). Because trial characteristics vary widely, what is needed is a collation of many studies to answer this question; however, sufficient information is not available to do so.

The other criterion is how the duration of follow-up influences the effect size. Many treatments for behavioral disorders have initial effects that later dissipate, which has been a major reason for the use of 6- or 12-month follow-ups (Hughes et al., 1992). Whether or not the effect size for treatments for smoking changes over time is unclear. Most meta-analyses indicate the effect size is similar across follow-ups of different durations, and one meta-analysis reported no statistically significant difference across follow-ups (Gould & Clum, 1993) (Table 2a and b). Other meta-analyses reported that the effect size diminished at long-term follow-up, and one showed this difference was statistically significant (Law & Tang, 1995).

Recommendations

Although some current meta-analyses suggest effect sizes, as measured by odds ratios, do not differ between early and later follow-up (which would suggest that shorter follow-ups should be sufficient), we recommend that the convention of using 6- or 12-month follow-ups be continued until a more definite conclusion is reached. Our group could not reach consensus on a preference for 6- vs. 12-month follow-ups. The major advantages of a 6-month follow-up are that several recent meta-analyses have used that length as the standard because of its high prevalence of use in studies (Hughes, 1996). A 6-month rather than a 12-month follow-up also allows for more rapid decisions on treatment efficacy. The major advantages of a 12-month follow-up are that longer follow-ups are associated more closely with life-long abstinence, are probably less likely to give false positive results, and may be more persuasive to scientists and non-scientists.

Table 2a. Meta-analyses that included quantitative reports of effect sizes for different follow-ups

Meta-analysis	Treatment	Effect size measure	Effect size (95% *CI*) 1 month[a]	3 month[a]	6 month[a]	12 month[a]
Cepeda-Benito, 1993	Nicotine gum	*d* (Intense)	0.40 (0.27–0.53)	–	–	0.28 (0.16–0.40)
		d (Brief)	0.19 (0.10–0.28)	–	–	0.04 (−0.05–0.12)
Fiore et al., 1994	Nicotine patch	*OR*	–	2.6 (2.2–3.0)	3.0	– (2.4–3.7)
Hughes, 1991	Nicotine patch + Beh tx	|diff|(Nic)	–	15	–	13
		|diff| (beh. tmt.)	–	15	–	12
Kottke et al., 1988	Multiple	|diff|	–	–	8.4 (5.6–11.2)	5.8 (3.2–8.4)
Law & Tang, 1995	Multiple	|diff|	0.25 (0.22–0.27)[c]	0.18 (0.15–0.21)[c]	0.14 (0.11–0.17)[c]	0.11 (0.08–0.14)
Lichtenstein, 1996	Telephone counseling	*OR*	–	–	1.3 (1.2–1.5)	1.2 (1.1–1.8)
Po, 1993	Nicotine patch	*OR*	3.1 (2.7–3.6)	–	–	2.3 (1.8–2.9)
Stapleton, 1998	Nicotine replacement	*OR* (relapse)[b]	0.9 (0.7–1.2)	0.9 (0.7–1.1)	–	–
White & Rampes, 1999	Acupuncture	*OR*	1.2 (1.0–1.5)	–	1.4 (0.9–2.1)	1.0 (0.7–1.4)

CI, confidence interval; *OR*, odds ratio; *d*, Probit change score; Intense, with intensive behavior therapy; Brief, with brief advice; |diff|, absolute difference in percentage abstinence; Nic, Nicotine gum effect size; Beh tx, Behavior therapy effect size; *RR*, Relative risk.
[a]When effect sizes were reported for short-term or long-term results, we assigned the result to the 1-, 3-, 6-, or 12-month column on the basis of the majority of the follow-up times in the data.
[b]Odds of relapse in active vs. control between this time and 12-month follow-up. If relapse *OR* does not differ from 1.0, *OR*s for shorter and longer follow-ups do not differ.
[c]Differs statistically from 12 months ($p < .05$).

Table 2b. Meta-analyses that reported text comments on effect sizes for different follow-ups

Meta-analysis	Treatment	Effect size	Comments
Gould & Clum, 1993	Self-help	–	"Not statistically different" between end of treatment and follow-up
Hughes, 1999	Nicotine gum	*OR, RR*, |diff|	"*OR* and *RR* was similar but |diff| decreased somewhat across follow-ups"
Lam et al., 1987	Nicotine gum	|diff|	No apparent difference in 1, 3, 6 and 12 months effect sizes in study figures
Silagy, 1994	Nicotine replacement	*OR*	"Fairly consistent" across 1, 3, 6, and 12 months

OR, odds ratio; *RR*, relative risk; |diff|, absolute difference in percentage abstinence.

Non-traditional measures of efficacy

The most commonly used non-traditional measures in tobacco clinical trials are those derived from an analysis of a plot of the relapse curve (i.e., survival analyses) (Koehler & McGovern, 1990). In addition, measures from event history analysis can model not only relapse but also recycling of smokers from relapse back to abstinence (Swan & Denk, 1987). These analyses provide much more information than examination of a single, arbitrarily defined point in time does (e.g., identification of high-risk periods and differentiation of causes of early vs. late relapse).

However, applying survival analysis to smoking entails potential problems. First, if survival curves use any smoking to define failure, it is unclear how to handle smokers who never quit. One way would be to count non-quitters as failures on Day 1. Another is to simply note that the analysis applies only to those who quit for a certain number of hours or days initially. Second, although one can think of survival analyses for a clinical trial as comparing two curves, the most commonly derived measure is the median time-to-relapse. The important issue is whether time-to-relapse is a relevant clinical outcome. A treatment could produce a greater median time-to-relapse, but

not a greater number of smokers abstinent at a long-term follow-up (see Fig. 4). Whether or not such an outcome indicates a treatment might be useful is debatable.

Another alternative measure is the total amount of smoking in a group (e.g., the cumulative number of cigarettes for all subjects in a group over a specified period). The rationale is that cross-sectional data indicate that the risks of smoking are highly dose-related to the total amount smoked (USDHHS, 1990). Thus if an experimental group of 100 subjects decreased its smoking from 2000 to 1500 cigarettes per day, but the control group did not change its total of 2000 cigarettes per day, one could conclude that the treatment reduced the risks of smoking. The major argument against this rationale is that many smokers smoke more intensely when they reduce their cigarettes per day, which would negate some of the health benefits of reduction (Scherer, 1999).

Another similar measure is one that focuses not on the amount of smoking, but on the total number of days smoking in a group. The major rationale here is that recent analyses suggest the risks of smoking, at least on a cross-sectional basis, are more highly related to duration of use than to number of cigarettes per day (Peto, 1996).

Another measure is the longest duration of abstinence during the follow-up. The rationale for this measure is that the longest duration of abstinence is one of the best, if not the best, predictor of future success at cessation (Gilpin, Pierce, & Farkas, 1997). If we assume that this association is, to some degree, causal, a treatment that produces a longer duration of abstinence should increase the probability of eventual life-long abstinence.

Other measures used in non-nicotine drug dependence trials appear less appropriate for nicotine dependence (e.g., number of days of heavy use, time to first heavy use period, severity or frequency of drug problems, abuse, dependence) (Breslin, Sobell, Sobell, & Sobell, 1997; Floyd, Monahan, Finney, & Morley, 1996).

Recommendation

We recommend that results from survival analyses or event history analyses be reported in addition to prolonged abstinence because they provide additional information to simple quit rates.

Summary of recommendations

The SRNT Measures Workgroups recommended that each subcommittee use the following Levels of Evidence ratings: A — multiple well-designed studies that show a consistent pattern of findings, B — some evidence to support the recommendation but scientific

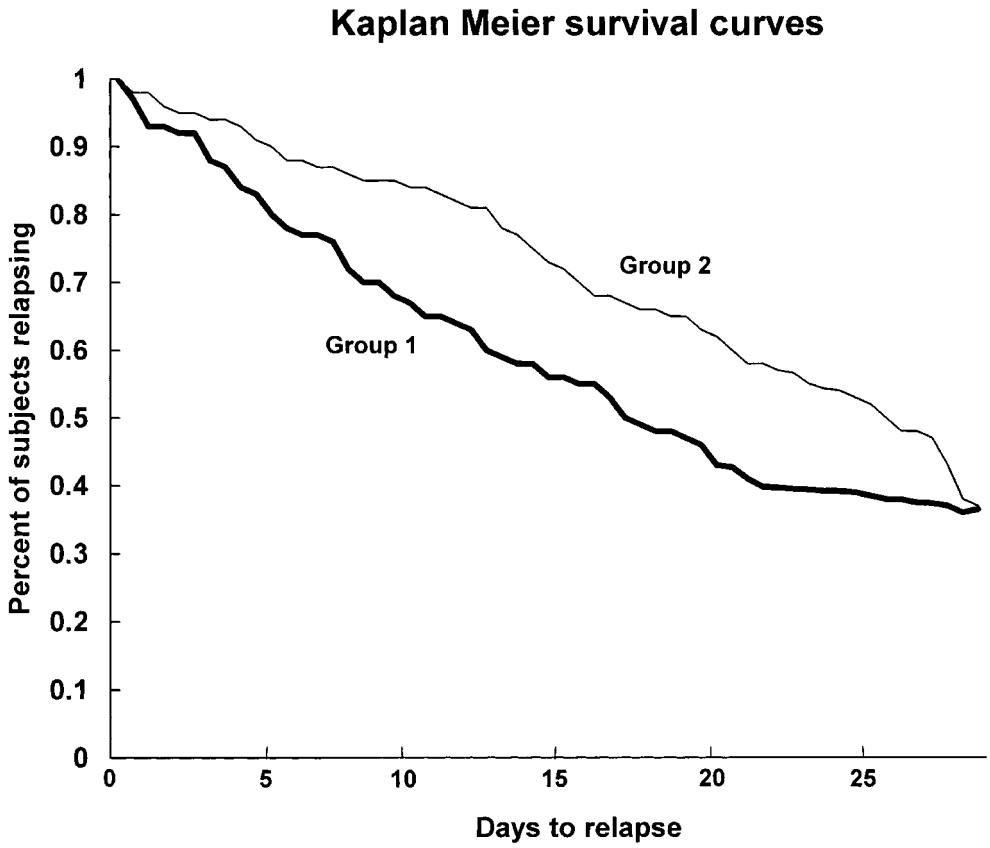

Figure 4. Curves that differ in time-to-relapse but not in final outcome.

support was not optimal, and C — the group achieved consensus, but in the absence of scientific support. Because of a lack of data, we have no A recommendations. *Thus, all of our recommendations should be considered tentative.* In fact, until definitive conclusions can be reached on the basis of empirical analyses, the best approach is to report multiple measures of abstinence for each study. However, if an author is pressed to reduce the number of outcomes reported, we suggest the recommendations shown in Table 3.

Future research. Meta-analyses should examine how much abstinence rates and effect sizes differ using point prevalence, vs. repeated point prevalence, vs. continuous, vs. prolonged abstinence; using different definitions of failure; and using short vs. long vs. very long follow-ups.

Future clinical trials should report several outcomes. Having multiple outcomes (e.g., both 6- and 12-month follow-ups) allows empirical determination of whether measures are equivalent or whether one measure is better than another and allows meta-analyses to examine the methodological issues outlined here. Table 4 proposes text for questions necessary to obtain the different abstinence outcomes. The text has not been tested, and thus empirical demonstration of the optimal wording for these questions is needed.

Second, because pressure from journal editors to limit the length of articles is a perceived barrier to displaying several outcomes, we recommend that relevant organizations (e.g., SRNT, the College on Problems of Drug Dependence, and the American Public Health Association) produce a white paper to be sent to the editors of journals that regularly publish clinical trials in smoking cessation (e.g. *Journal of the American Medical Association*, *Addiction*) and related organizations (e.g., the Cochrane Collaboration, the Society for Clinical Trials) urging them to recommend reporting outcomes in several formats. We believe journals affiliated with scientific organizations (e.g. *Nicotine & Tobacco Research* and *Drug and Alcohol Dependence*) should lead this process.

Third, clinical investigators should publish non-traditional outcome measures and compare them with traditional outcomes. Our effort was begun, in large part, because of fears of increasing rigidity in the measurement of clinical trial outcomes. *Thus, we do not want our recommendations to be applied as rigid standards.* Rather, we hope that by raising issues, we have encouraged trialists not only to report standard outcomes, but also to be more flexible, curious and adventurous in examining ways to measure abstinence.

Fourth, some of our best data on outcomes come from comments in meta-analyses (Fiore et al., 1994; Fisher et al., 1990; Gould & Clum, 1993; Law & Tang,

Table 3. Recommended minimum reporting of abstinence measures in clinical trials

For all trials
1. Report prolonged abstinence as the primary outcome measure, failure defined either as 7 consecutive days of smoking or smoking on at least 1 day on each of 2 consecutive weeks (B).[a]
2. Use the term *continuous abstinence* only to refer to prolonged abstinence that has no grace period (B).
3. Do not use the term *sustained abstinence* (B).
4. Include non-cigarette tobacco use in definitions of failure (B).
5. Do not include non-tobacco nicotine use (e.g., NRT) in definitions of failure (B).
6. When feasible, report survival or event history analysis outcomes (B)

Aid-to-cessation and proof of concept trials
1. Tie all follow-ups to the experimenter or subject-determined quit date (B).
2. Use a grace period for prolonged abstinence of 2 weeks unless there is an *a priori* rationale for a different duration. If there is *a priori* rationale, use no grace period (B).
3. Report 6- and/or 12-month follow-ups (B).
4. Report point prevalence as a secondary measure in which failure is defined as any smoking in the 7 days before the follow-up (B) If CO, SCN or some other non-cotinine measure is used to verify, also include a window that is biochemically verifiable (C).
5. Report point prevalence that uses a 30-day window (C).

Cessation induction trials
1. Tie all follow-ups to the onset of treatment (B).
2. State the grace period for prolonged abstinence a priori based on the anticipated duration of treatment dissemination or other factors (B).
3. Report 6- and/or 12-month follow-ups (B).
4. Report point prevalence as a secondary measure in which failure is defined as any smoking in the 7 days before the follow-up (B). If CO, SCN or some other non-cotinine measure is used to verify, also include a window that is biochemically verifiable (C).
5. Report point prevalence, which also uses a 30-day window (C).

[a]A, multiple well-designed studies that showed a consistent pattern of findings; B, some evidence to support the recommendation, but scientific support was not optimal; and C, the group achieved consensus, but in the absence of scientific support.
NRT, nicotine replacement therapy; CO, carbon monoxide; SCN, thiocyanate.

Table 4. Questions necessary to obtain abstinence measures[a]

1. Since [the end of the grace period] have you ever smoked at least a part of a cigarette on each of 7 consecutive days? After [the end of the grace period] have you smoked any in each of 2 consecutive weeks? (Prolonged abstinence)
2. What was the first day of that 7-day or 2-week period you smoked? (Survival analysis)
3. Since [the quit date for aid-to-cessation trials or beginning of treatment, or end of treatment for cessation induction trials] have you smoked at least a part of a cigarette on each of 7 consecutive days, or have you smoked any in each of 2 consecutive weeks? (Continuous abstinence) Have you smoked at least part of a cigarette in the last 7 days? If not, have you smoked at least part of a cigarette in last 30 days? (Point prevalence and repeated point prevalence)

1995; Silagy et al., 1994a, b; Tang, Law & Wald, 1994). We urge those undertaking meta-analyses to report in much more detail on methodological issues.

Fifth, we encourage trialists to combine and compare their data sets to explore methodological issues. For example, some of the best existing data on outcomes measures come from a collaboration of several trialists studying self-quitting (Cohen et al., 1989). Similarly, some pharmaceutical companies have used similar outcomes across several studies and could combine their data sets to examine methodological issues.

Acknowledgments

We thank the more than 20 SRNT members who provided comments on this manuscript. The writing of the manuscript was supported by Research Scientist Development Award 00109 to John R. Hughes from the National Institute on Drug Abuse.

References

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders: Fourth Edition*. Washington, DC: Author.

Anton, R. F., Moak, D. H., Waid, L. R., Latham, P. K., Malcolm, R. J., & Dias, J. K. (1999). Naltrexone and cognitive behavioral therapy for the treatment of outpatient alcoholics: Results of a placebo-controlled trial. *American Journal of Psychiatry*, *156,* 1758–1764.

Becona, E., & Vazquez, F. L. (1998). The course of relapse across 36 months for smokers from a smoking cessation program. *Psychological Reports*, *82,* 143–146.

Benowitz, N. L. (1998). *Nicotine Safety and Toxicity*. New York: Oxford University Press.

Benowitz, N., Ahijevych, K., Hall, S., Hansson, A., Henningfield, J. E., Hurt, R. D., Jacob, P., Jarvis, M. J., LeHouezec, J., Lichtenstein, E., Tsoh, J., & Velicer, W. (2002). Biochemical verification of tobacco use and cessation. *Nicotine and Tobacco Research*, *4,* 149–159.

Borland, R. (1994). Population estimates of occasional smoking among self-described smokers and non-smokers in Victoria, Australia. *Tobacco Control*, *3,* 37–40.

Brandon, T. H., Lazev, A. B., & Juliano, L. M. (1998). Very delayed smoking relapse warrants research attention. *Psychological Reports*, *83,* 72–74.

Breslin, F. C., Sobell, S. L., Sobell, L. C., & Sobell, M. B. (1997). Alcohol treatment outcome methodology: State of the art 1989–1993. *Addictive Behaviors*, *22,* 145–155.

Cepeda-Benito, A. (1993). Meta-analytical review of the efficacy of nicotine chewing gum in smoking treatment programs. *Journal of Consulting and Clinical Psychology*, *61,* 822–830.

Cohen, S., Lichtenstein, E., Prochaska, J. O., Rossi, J. S., Gritz, E. R., Carr, C. R., Orleans, C. T., Schoenbach, V. J., Biener, L., Abrams, D., DiClemente, C. C., Curry, S., Marlatt, G. A., Cumming, K. M., Emont, S. L., Giovino, G. A., & Ossip-Klein, D. J. (1989). Debunking myths about self-quitting. *American Psychologist*, *44,* 1355–1365.

College on Problems of Drug Dependence (1999). Consensus statement on evaluation of outcome of pharmacotherapy for substance abuse/dependence. Retrieved from views.vcu.edu/cpdd/reports/nida_cpdd_report.pdf on October 29, 2002.

Covey, L. S., & Glassman, A. H. (1991). A meta-analysis of double-blind placebo-controlled trials of clonidine for smoking cessation. *British Journal of Addiction*, *86,* 991–998.

Fiore, M. C., Bailey, W. C., Cohen, S. J., Dorfman, S. F., Goldstein, M. G., Gritz, E. R., Heyman, R. B., Jaen, C. R., Kottke, T., Lando, H. S., Mecklenburg, R. E., Mullen, P. D., Nett, L. M., Robinson, L., Stizer, M. L., Tommasello, A. C., Villejo, L., & Wewers, M. E. (2000). *Treating Tobacco Use and Dependence. Clinical Practice Guideline*. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service.

Fiore, M. C., Smith, S. S., Jorenby, D. E., & Baker, T. B. (1994). The effectiveness of the nicotine patch for smoking cessation: A meta-analysis. *Journal of the American Medical Association*, *271,* 1940–1946.

Fisher, K. J., Glasgow, R. E., & Terborg, J. R. (1990). Work site smoking cessation: A meta-analysis of long-term quit rates from controlled studies. *Journal of Occupational Medicine*, *32,* 429–439.

Floyd, A. S., Monahan, S. C., Finney, J. W., & Morley, J. A. (1996). Alcoholism treatment outcome studies, 1980–1992: The nature of the research. *Addictive Behaviors*, *21,* 413–428.

Gerlach, K. K., Cummings, K. M., Hyland, A., Gilpin, E. A., Johnson, M. D., & Pierce, J. P. (1998). Trends in cigar consumption and smoking prevalence. In D. M. Burns, D. Hoffmann, K. M. Cummings, & R.H. Amacher (Eds.), *Smoking and Tobacco Control Monograph 9: Cigars Health Effects and Trends* (pp. 21–53). Bethesda, MD: U.S. Department of Health and Human Services.

Gilpin, E. A., Pierce, J. P., & Farkas, A. J. (1997). Duration of smoking abstinence and success in quitting. *Journal of the National Cancer Institute*, *89,* 572–576.

Gilpin, E., Cavin, S. W., & Pierce, J. P. (1997). Adult smokers who do not smoke daily. *Addiction*, *92,* 473–480.

Gould, R. A., & Clum, G. A. (1993). A meta-analysis of self-help treatment approaches. *Clinical Psychology Review*, *13,* 169–186.

Gourlay, S. G., & Benowitz, N. L. (1995). Is clonidine an effective smoking cessation therapy? *Drugs*, *50,* 197–207.

Gourlay, S. G., & McNeil, J. J. (1990). Antismoking products. *Medical Journal of Australia*, *153,* 699–707.

Grabowski, J., & Bell, C. S. (1983). *Measurement in the analysis and treatment of smoking behavior. NIDA Research Monograph 48* (pp. 83–1285). Washington, DC: U.S. Government Printing Office.

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

Henningfield, J. E., Ahluwalia, J. S., Dalack, G. W., Burns, D., & Etter, J. F. et al. (2000, February). *Treatment outcome measures for the studies of reduction of cigarette smoking*. Paper presented at Annual Meeting of the Society for Research on Nicotine and Tobacco, Arlington, VA.

Hughes, J. R. (1991). Combined psychological and nicotine gum treatment for smoking: A critical review. *Journal of Substance Abuse*, *3,* 337–350.

Hughes, J. R. (1996). Pharmacotherapy of nicotine dependence. In C. R. Schuster & M. J. Kuhar (Eds.), *Pharmacological Aspects of Drug Dependence: Toward an Integrative Neurobehavioral Approach, Handbook of Experimental Pharmacology Series* (pp. 599–626). New York: Springer-Verlag.

Hughes, J. R. (1998). Dependence on and abuse of nicotine replacement medications: An update. In N. L. Benowitz (Ed.), *Nicotine Safety & Toxicity* (pp. 147–157). New York: Oxford Press.

Hughes, J. R. (1999, March). *Is long-term follow-up necessary in clinical trials?* Paper presented at annual meeting of Society for Research on Nicotine and Tobacco, New Orleans.

Hughes, J. R., Gulliver, S. B., Fenwick, J. W., Cruser, K., Valliere, W. A., Pepper, S. L., Shea, P. J., & Solomon, L. J. (1992). Smoking cessation among self-quitters. *Health Psychology*, *11,* 331–334.

Jamerson, B. D., Nides, M., Jorenby, D. E., Donahue, R., Garrett, P., Johnston, J. A., Fiore, M. C., Rennard, S. I., & Leischow, S. J. (2001). Late-term smoking cessation despite initial failure: An evaluation of bupropion sustained release, nicotine patch, combination therapy, and placebo. *Clinical Therapeutics*, *23,* 744–752.

Jarvis, M., West, R., Tunstall-Pedoe, H., & Vesley, C. (1984). An evaluation of the intervention against smoking in the Multiple Risk Factor Intervention Trial. *Preventive Medicine*, *13,* 501–509.

Kenford, S. L., Fiore, M. C., Jorenby, D. E., Smith, S. S., Wetter, D., & Baker, T. B. (1994). Predicting smoking cessation: Who will quit with and without the nicotine patch. *Journal of the American Medical Association*, *271,* 589–594.

Killen, J. D., Fortmann, S. P., Newman, B., & Varady, A. (1990). Evaluation of a treatment approach combining nicotine gum with self-guided behavioral treatments for smoking relapse prevention. *Journal of Consulting and Clinical Psychology*, *58,* 85–92.

Koehler, K. J., & McGovern, P. G. (1990). An application of the LFP

survival model to smoking cessation data. *Statistics in Medicine*, *9,* 409–421.

Kottke, T. E., Battista, R. N., DeFriese, G. H., & Brekke, M. L. (1988). Attributes of successful smoking cessation interventions in medical practice: A meta-analysis of 39 controlled trials. *Journal of the American Medical Association*, *259,* 2882–2889.

Lam, W. L., Sze, P. C., Sacks, H. S., & Chalmer, T. C. (1987). Meta-analysis of randomized controlled trials of nicotine chewing gum. *Lancet*, *2,* 27–29.

Lando, H. A. (1983). Data collection and questionnaire design: Smoking cessation in adults. *Measurement in the Analysis and Treatment of Smoking Behavior: NIDA Research Monograph Series 48* (pp. 74–89). Rockville, MD: U.S. Government Printing Office.

Law, M., & Tang, J. L. (1995). An analysis of the effectiveness of interventions intended to help people stop smoking. *Archives of Internal Medicine*, *155,* 1933–1941.

Lichtenstein, E., Glasgow, R. E., Lando, H. A., & Ossip-Klein, D. J. (1996). Telephone counseling for smoking cessation: Rationale and review of evidence. *Health Education Research*, *11,* 243–257.

Mermelstein, R. J., Colby, S. M., Patten, C., Prokhorov, A., Brown, R., Myers, M., Adelman, W., Hudmon, K., & McDonald, P. (2002). Methodological issues in measuring treatment outcome in adolescent smoking cessation studies. *Nicotine and Tobacco Research*, *4,* 395–403.

Midanik, L. T., Polen, M. R., Hunkeler, E. M., Tekawa, I. S., & Soghikian, K. (1985). Methodologic issues in evaluating stop smoking programs. *American Journal of Public Health*, *75,* 634–638.

Miller, W. R. (1996). What is a relapse? Fifty ways to leave the wagon. *Addiction*, *91,* S15–S27.

Ockene, J. K., Kristeller, J., Pbert, L., Hebert, J. R., Luippold, R., Goldberg, R. J., Landon, J., & Kalan, K. (1994). The physician-delivered smoking intervention project: Can short-term interventions produce long-term effects for a general outpatient population? *Health Psychology*, *13,* 278–281.

Ockene, J. K., Emmons, K. M., Mermelstein, R. J., Perkins, K. A., Bonollo, D. S., Voorhees, C. C., & Hollis, J. F. (2000). Relapse and maintenance issues for smoking cessation. *Health Psychology*, *19,* 17–31.

Ossip-Klein, D. J., Bigelow, G., Parker, S. R., Curry, S., Hall, S., & Kirkland, S. (1986). Classification and assessment of smoking behavior. *Health Psychology*, *5,* 3–11.

Peto, R. (1996). Influence of dose and duration of smoking on lung cancer rates. In D. Zaridze & R. Peto (Eds.), *Tobacco: A Major International Health Hazard* (pp. 23–33). Lyon: ARC Scientific Publications No. 74.

Po, A. L. W. (1993). Transdermal nicotine in smoking cessation. *European Journal of Clinical Pharmacology*, *45,* 519–528.

Prochaska, J. O., DiClemente, C. C., Velicer, W. F., & Rossi, J. S. (1993). Standardized, individualized, interactive and personalized self-help programs for smoking cessation. *Health Psychology*, *12,* 399–405.

Ramstrom, L. M. (1987). The WHO guidelines for standardization of smoking surveys. In M. Aoki, S. Hisamichi, & S. Tominaga (Eds.), *Smoking and Health* (pp. 715–717). Amsterdam: Elsevier Science Publishers.

Richmond, R. L. (1997). A comparison of measures used to assess effectiveness of the transdermal nicotine patch at 1 year. *Addictive Behaviors*, *22,* 753–757.

Russell, M. A. H., Merriman, R., Stapleton, J., & Taylor, W. (1983). Effect of nicotine chewing gum as an adjunct to general practitioners' advice against smoking. *British Medical Journal*, *287,* 1782–1785.

Scherer, G. (1999). Smoking behaviour and compensation: a review of the literature. *Pschopharmacology*, *145,* 1–20.

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,* 1671–1682.

Shanks, T. G., & Burns, D. M. (1998). Disease consequences of cigar smoking. In D. M. Burns, D. Hoffmann, K. M. Cummings, & R.H. Amacher (Eds.), *Smoking and Tobacco Control Monograph 9: Cigars Health Effects and Trends* (pp. 105–160). Bethesda, MD: U.S. Department of Health and Human Services.

Shiffman, S., West, R. J. & Gilbert, D. (2000, February). *The assessment of tobacco craving and withdrawal in smoking cessation trials*. Paper presented at the Annual Meeting of the Society for Research on Nicotine and Tobacco, Arlington, VA.

Shipley, R. H., Rosen, T. J., & Williams, C. (1982). Measurement of smoking: Surveys and some recommendations. *Addictive Behaviors*, *7,* 299–302.

Silagy, C., Mant, D., Fowler, G., & Lodge, M. (1994). Meta-analysis on efficacy of nicotine replacement therapies in smoking cessation. *Lancet*, *343,* 139–142.

Silagy, C., Mant, D., Fowler, G., & Lodge, M. (1994). The effectiveness of nicotine replacement therapies in smoking cessation. *Online Journal of Current Clinical Trials*, *3* (14 January 1994), document 113.

Stapleton, J. (1998). Cigarette smoking prevalence, cessation and relapse. *Statistical Methods in Medical Research*, *7,* 187–203.

Sutherland, G., Stapleton, J. A., Russell, M. A. H., Jarvis, M. J., Hajek, P., Belcher, M., & Feyerabend, C. (1992). Randomised controlled trial of nasal nicotine spray in smoking cessation. *Lancet*, *340,* 324–329.

Swan, G. E., & Denk, C. E. (1987). Dynamic models for the maintenance of smoking cessation: Event history analysis of late relapse. *Journal of Behavioral Medicine*, *10,* 527–554.

Tang, J. L., Law, M., & Wald, N. (1994). How effective is nicotine replacement therapy in helping people to stop smoking? *British Medical Journal*, *308,* 21–25.

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,* 311–315.

U.S. Department of Health and Human Services. (1990). *Health Benefits of Smoking Cessation. A Report of the U.S. Surgeon General*. Rockville, MD: National Institute on Drug Abuse.

U.S. Food and Drug Administration. (1995). *Transcript of the Joint Meeting of the Nonprescription Drugs Advisory Committee and the Drug Abuse Advisory Committee of the Food and Drug Administration.* Rockville, MD: Author.

Velicer, W. F., & Prochaska, J. O. (in press). A comparison of four self-report smoking cessation outcome measures. *Addictive Behaviours*.

Velicer, W. F., Prochaska, J. O., Rossi, J. S., & Snow, M. G. (1992). Assessing outcome in smoking cessation studies. *Psychological Bulletin*, *111,* 23–41.

White, A. R., Rampes, H., & Ernst, E. (1999). Acupuncture for smoking cessation. *The Cochrane Library*, *4,* 1–15.