Published in final edited form as:

Clin Trials. 2010 June ; 7(3): 286–298. doi:10.1177/1740774510367811.

A tutorial on principal stratification-based sensitivity analysis: Application to smoking cessation studies

Brian L. Egleston, PhD¹, Karen L. Cropsey, PsyD², Amy B. Lazev, PhD³, and Carolyn J. Heckman, PhD³

¹Biostatistics and Bioinformatics Facility, Fox Chase Cancer Center, Philadelphia, PA, USA

²Department of Psychiatry and Behavioral Neurobiology, University of Alabama School of Medicine, Birmingham, AL, USA

³Cancer Prevention and Control Program, Fox Chase Cancer Center, Philadelphia, PA, USA

Abstract

Background—One problem with assessing effects of smoking cessation interventions on withdrawal symptoms is that symptoms are affected by whether participants abstain from smoking during trials. Those who enter a randomized trial but do not change smoking behavior might not experience withdrawal related symptoms.

Purpose—We present a tutorial of how one can use a principal stratification sensitivity analysis to account for abstinence in the estimation of smoking cessation intervention effects. The paper is intended to introduce researchers to principal stratification and describe how they might implement the methods.

Methods—We provide a hypothetical example that demonstrates why estimating effects within observed abstinence groups is problematic. We demonstrate how estimation of effects within groups defined by potential abstinence that an individual would have in either arm of a study can provide meaningful inferences. We describe a sensitivity analysis method to estimate such effects, and use it to investigate effects of a combined behavioral and nicotine replacement therapy intervention on withdrawal symptoms in a female prisoner population.

Results—Overall, the intervention was found to reduce withdrawal symptoms but the effect was not statistically significant in the group that was observed to abstain. More importantly, the intervention was found to be highly effective in the group that would abstain regardless of intervention assignment. The effectiveness of the intervention in other potential abstinence strata depends on the sensitivity analysis assumptions.

Limitations—We make assumptions to narrow the range of our sensitivity parameter estimates. While appropriate in this situation, such assumptions might not be plausible in all situations.

Conclusions—A principal stratification sensitivity analysis provides a meaningful method of accounting for abstinence effects in the evaluation of smoking cessation interventions on withdrawal symptoms. Smoking researchers have previously recommended analyses in subgroups defined by observed abstinence status in the evaluation of smoking cessation interventions. We believe that principal stratification analyses should replace such analyses as the preferred means of accounting for post-randomization abstinence effects in the evaluation of smoking cessation programs.

Introduction

We present a tutorial of how principal stratification [1] can be used to simplify controlling for the impact of post-randomization intermediate variables on outcomes in clinical trials. As the name implies, a post-randomization intermediate variable is a variable measured post-randomization that is affected by randomization assignment and also affects other more distal post-randomization outcomes. A common example of a post-randomization variable is compliance. A patient's compliance with treatment is often affected by randomization assignment. Subsequently, a patient's outcome will be affected by both randomization assignment and level of compliance with the assigned treatment.

Historically, intermediate variables have been examined as mediator variables. A graphical representation of the pathway would be that shown in figure 1. We have a direct effect of treatment assignment on an outcome as represented by arrow c. We also have an indirect pathway of treatment assignment on an outcome that is through treatment assignment's impact on the intermediate variable and the intermediate variable's impact on the more distal outcome (defined by arrows a and b). Authors such as Judea Pearl have formalized a language for such graphical representations of pathways.[2]

Principal stratification presents a relatively new method of examining the impact of intermediate variables on outcomes. It differs from the traditional path and structural equation models that have been used to investigate direct and indirect effects.[2] In a two arm trial, principal stratification refers to the subgrouping of individuals by the potential outcomes that they would have had if they had been assigned to each arm of the study. In practice, this often requires estimating the same individual's intermediate and distal outcomes had she been assigned to intervention and had she been assigned to control. This helps us better understand the effects of intermediate and distal outcomes on different groups of individuals. Defining meaningful causal effects using a principal stratification framework has become increasingly common.[3][4][5][6][7][8][9][10][11][12][13][14][15]

This paper provides conceptual details concerning principal stratification and a data example of the approach. The paper is organized as follows. We frame the issue using a hypothetical example in the following section. Next, we present a data example in which we discuss how one can estimate the effects of interest. We report the results of our data example, and then provide a discussion of our findings. An appendix gives technical details of the approach, and a glossary of commonly used terms in the paper is provided.

Framing of the Issue

To provide a more concrete example, we present principal stratification methodology that accounts for the complicating impact of abstinence status on withdrawal symptoms in smoking cessation intervention studies. Here we assume that the intervention comprises nicotine replacement therapy with counseling while the control condition consists of no active treatment. One problem with assessing the effect of an intervention on withdrawal symptoms is that symptoms are affected by whether a person makes an attempt to quit or actually abstains from smoking during the intervention. Intuitively, those who enter a randomized trial but do not change their smoking behavior after enrollment in the trial might not experience any withdrawal related symptoms. In a randomized trial, it is likely that smoking behavior changes will differ between two arms of a study. If many of those in the intervention arm of a study abstain from smoking, but few of those in the control arm make similar attempts, those in the intervention arm might report more withdrawal symptoms. Investigators might naively conclude that the intervention increased withdrawal symptoms when different abstinence rates actually drove the relationship.

To demonstrate how abstention can affect inferences we present a hypothetical example and a framework for investigations. The potential outcomes framework developed by Neyman, [16] Rubin,[17] and Holland,[18] provides a meaningful way to characterize the difficulty in assessing intervention effects on withdrawal symptoms. Under the potential outcomes framework, we conceptualize that individuals in a study with two randomization arms have two potential outcomes: one outcome under the control condition and one outcome under the intervention condition. An assumption we are making is that there are well defined intervention and control conditions; that is, the treatments offered to individuals within randomization arms do not change depending on when or how individuals are assigned to arms (for more details, see discussions of consistency in Pearl [2] and stable-unit-treatment - value assumptions in Rubin [19]).

Using the potential outcomes framework, we can group individuals based on their potential abstention status. Thus, we have 1) a group that abstains regardless of whether they are assigned to the intervention or the control condition, 2) a group that abstains only when assigned to the intervention (but does not abstain when assigned to the control), 3) a group that would not abstain regardless of whether they are assigned to the intervention or control condition, and 4) a group that abstains only when assigned to the control condition (but does not abstain when assigned to the intervention). Each of these four groups has been termed a "principal stratum" by Frangakis and Rubin.[1]

For purposes of this example, we will assume that there is no one in the fourth group. A number of authors have proposed such a strict "monotonicity" assumption eliminating the fourth group for the identification of principal strata causal effects.[7][8][9][15] Such an assumption is appropriate in this context as there is strong evidence that nicotine replacement therapy combined with counseling is highly effective.[20][21] Hence, it is not realistic to presume that a person who quits in a control condition that does not include any active intervention would be induced to not quit when given nicotine replacement therapy with counseling. Further, as will be described later, there is substantial reduction in the range of point estimates if such an assumption can be made validly.

In table 1, we present a hypothetical example of the percentage with and without withdrawal symptoms on a dichotomous symptom scale by principal stratum and intervention assignment status. In the table, principal stratum assignment is not related to intervention assignment. That is, within each principal stratum, there are equal numbers of individuals assigned to the control arm and the intervention arm. We would generally expect this to be the case on average since individuals cannot be randomized to interventions within the unknown principal strata.

Also shown in table 1, the intervention decreases symptoms among those who always abstain (3% out of 10% among control arm, versus 2% out of 10% in intervention arm). This group abstains regardless of condition, but the intervention (e.g. NRT) helps reduce their nicotine withdrawal symptoms. The intervention worsens withdrawal symptoms in those who can abstain with the intervention only (3% out of 15% in control arm versus 6% out of 15% in intervention arm). In this group, because people become abstinent from nicotine in the intervention but not the control condition, being assigned to the intervention increases the severity of withdrawal symptoms. Finally, the intervention has no effect in the group that cannot abstain regardless of intervention assignment (we call this the group that never abstains). We believe that some of those who do not abstain might still experience withdrawal symptoms due to attempts to reduce the amount of smoking.

In table 2, we present the data that a researcher would actually observe. The researcher only knows the outcomes in the group to which an individual is assigned, not the outcomes that

the individual would have had in the other assignment group. Among the observed abstainers in the sample, 30% ($3\% / 10\% = 30\%$) in the control arm report withdrawal symptoms compared with 32% ($8\% / 25\% = 32\%$) in the intervention arm. Among observed non-abstainers, 20% ($8\% / 40\% = 20\%$) of those in the control group and 20% ($5\% / 25\% = 20\%$) in the intervention group report withdrawal symptoms. One might conclude that the intervention worsens withdrawal symptoms among those who are able to abstain during the study, but has no effect in those who never abstain during the study. However, we see from table 1 that such inferences differ by potential abstention status. In this paper, we propose to estimate the effects of a smoking cessation program on withdrawal symptoms in groups defined by potential abstention status as defined in table 1.

This framing of the problem illustrates why it is problematic to use the methods of Shiffman and colleagues [22] who advocated comparing intervention effects on withdrawal symptoms among multiple groups defined by observed abstinence status. Those who are observed to quit in the intervention arm are a mixture of those who would always abstain and those who would only abstain with intervention. Meanwhile, the group who would be observed to quit in the control arm consists only of those who would always abstain. Hence, individuals in the control group who abstained have different characteristics from individuals in the intervention group who abstained.

Data Example

Our data example concerns assessment of the impact of smoking cessation interventions on withdrawal symptoms. Interventions that reduce symptoms might improve quality of life during the quitting process and lead to higher cessation rates. Withdrawal symptoms include: craving, irritability, depression, restlessness, sleep disturbance, difficulty concentrating, increased appetite, weight gain, and anxiety.[22] Targeting amelioration of withdrawal symptoms is an important goal in aiding cessation since withdrawal symptoms, particularly craving, are associated with smoking relapse. While withdrawal patterns, time course, and severity may vary between individuals,[23][24] the vast majority of smokers experience withdrawal upon quitting smoking[25] with withdrawal symptoms being more severe for those who are more dependent on nicotine.[26]

We apply the methods to a randomized waitlist control smoking cessation intervention study on withdrawal symptoms in a female prisoner group from the southeastern U.S. from 2004 to 2006. As described by Cropsey and colleagues,[20] the intervention consisted of a behavioral component and a pharmacologic component. The intervention had a strong effect on increasing quit rates among study participants.

Intervention Description

The behavioral intervention was based on “Mood Management Training to Prevent Smoking Relapse”.[27] The 10-session group intervention was modified to conform to the prison environment and included typical examples of smoking triggers and coping strategies that were appropriate for a correctional environment.

All intervention participants received NicoDerm CQ® patches during week 3 of the intervention following the manufacturer’s suggested dosing. Participants who smoked more than 10 cigarettes per day (cpd) received 21 mg patches for 6 weeks, 14 mg for 2 weeks and 7 mg patches for 2 weeks for a total of 10 weeks of nicotine replacement (extending 2 weeks beyond the group intervention). Participants who smoked 10 or fewer cpd received 14 mg patches for 6 weeks and 7 mg patches for 2 weeks, for 8 weeks of nicotine replacement. Side effects were monitored and patches were distributed at weekly group sessions. Participants

were asked to make a quit attempt between weeks 3 and 4, immediately after receiving the first supply of nicotine replacement patches.

For purposes of this study, abstinence was defined as carbon monoxide (CO) exhalation of 2 parts per million (ppm) or less. This stringent cut-off is consistent with the carbon monoxide (CO) cut-off chosen by Cropsey and colleagues.[28] Point prevalence was used as a measure of abstinence as this would capture people who had recently started to abstain and hence might be near the peak of their withdrawal symptoms. Withdrawal symptoms were measured using the Minnesota Withdrawal Scale[29] both at baseline and follow-up.

We examined the relationship of the intervention with the first assessment of withdrawal symptoms after distribution of nicotine replacement therapy. Because control assessments were not as frequent as intervention assessments, the timing of evaluations differed. For those in the intervention arm, we used the first data available after distribution of the nicotine patch (week four for all but four participants whose first assessment was week five). For those in the control group, we used the first available check-in data after the baseline assessment (mean 10.6 weeks, standard deviation 10.0).

The original study used a wait-list control design; those assigned to the control group eventually received the smoking cessation program. For demonstration purposes, we only used control group data for those who were assigned initially to the wait-list control group to avoid correlated response data (having withdrawal symptom data under both control and intervention conditions among wait-list control participants would lead to correlated data). The intervention group in this paper consists of those initially assigned to the smoking cessation program.

Estimating Principal Stratum Causal Effects—To estimate effects within potential abstinence strata described in table 1, we adapted the potential outcomes paradigm that has become well known in the statistics literature. For example, Zhang and Rubin,[5] Hayden and colleagues,[6] and Egleston and colleagues[7][8] use the potential outcomes framework to investigate non-mortality outcomes when death is a significant competing risk. Similarly, Gilbert and colleagues[9] use the methods for investigation of HIV vaccine effects on viral loads, and Roy and colleagues[10] describe their use in assessing the impact of compliance on smoking cessation interventions. The estimators in each of these investigations fall under a generalizable framework of estimators developed to estimate effects within "principal strata"[1] defined by potential outcomes. Gallop and colleagues[11] and Jo[12] describe the use of principal stratification to investigate mediation, while Bellamy and colleagues[30] provide an overview of causal modeling in general.

For this work, we adapted the methods of Egleston and colleagues[7] to estimate the effects within principal strata. We contrast these estimates with the estimates in the observed abstainers. We also compared our estimates of the effect in the group that always abstains to the non-parametric bounding estimates of Zhang and Rubin[5], whose methods are similar to those of Manski and colleagues. [2][31][32]

Since we cannot directly examine the data and identify to which subgroup individuals belong, we need to make untestable but simplifying assumptions.[7] One key assumption is that the intervention does not decrease one's abstinence from smoking prior to evaluation. In the context of our example, the assumption assumes that the group that abstains only when assigned to the control condition (but not when assigned to the intervention condition) does not exist. As described in the framing of the problem section, this assumption is appropriate as there is strong evidence that an intervention with nicotine replacement therapy and

counseling will not decrease an individual's quitting compared to a control condition without any intervention.[20] [21]

We also use two sensitivity parameters that account for the lack of ability to identify groups defined by potential outcomes. Namely, let u and v define sensitivity parameters

$$u = \frac{\text{Average intervention withdrawal score if abstains with intervention only}}{\text{Average intervention withdrawal score if always abstains}}$$

and

$$v = \frac{\text{Average control withdrawal score if never abstains}}{\text{Average control withdrawal score if abstains with intervention only}}$$

Intuitively, the need for the sensitivity parameter u is related to the fact that we cannot separate in the intervention arm those who always abstain from those who abstain when assigned to intervention only (see table 2, second row of data). Similarly, we need v since we cannot separate those in the control arm who abstain when assigned to the intervention only from those who never abstain (see table 2, third row of data). Because we cannot estimate u and v using the observed data, we range over plausible values of u and v in a sensitivity analysis.

We use ratios for the sensitivity parameters as they are readily interpretable by investigators. They allow investigators to state the relationship on a relative rather than additive scale. An additive scale would necessitate speculating about differences in average symptom scores between groups (e.g., the differences could be up to three points), while a relative scale only requires speculation of proportional differences (e.g. one average is up to twice as large as the other). While the sensitivity parameters are not defined if the denominators are zero, this is not an issue as the withdrawal symptom scale is non-negative with very few zeros (2 out of 522 withdrawal symptom scores were equal to zero). Hence, the subgroup averages of interest will be greater than zero.

Since it is plausible that withdrawal symptoms generally worsen as people are less likely to abstain, we believe that u and v are both greater than one. In figures, we bound the sensitivity parameters at three since the ambiguity of the intervention effect's statistical significance with respect to the sensitivity parameters mostly falls within the range of one to three. The general trend of the intervention becomes more pronounced as the sensitivity parameters increase beyond three.

Finally, we developed multiple linear regression models for withdrawal symptoms and multiple logistic models for abstinence. We included the potential confounders described in the "Other Analytic Details" section below, but we also included abstinence status as a covariate in the symptoms model. Including confounders was not absolutely necessary since this was a randomized trial. However, their inclusion did assist in removing residual confounding due to chance imbalances between study arms or loss to follow-up. The utility of including confounders is discussed in more detail by others.[6][7][13][30]

More specifically, in order to estimate causal effects, we estimated multiple linear regression models separately for those in the control and intervention conditions. We then used the covariates from all participants to estimate four withdrawal symptom outcomes among the entire sample. First, we fixed the abstinence indicator to one (indicating abstinence) for everyone and estimated outcomes in the control and in the intervention models.

Notationally, let S_{CAi} and S_{TAi} denote the estimated respective control and intervention symptom scores for individual i as if the individual was an abstainer. Next, we fixed the abstinence indicator to zero (indicating no abstinence) for everyone and estimated outcomes in the control and intervention models. Let S_{CNI} and S_{TNI} indicate the estimated control and intervention symptom scores for individual i estimated as if there were no abstainers. We obtained S_{CAi} , S_{TAi} , S_{CNI} , and S_{TNI} for everyone in the sample.

Next, we fit separate abstinence logistic models for the control and intervention arms. We used these models to estimate abstinence outcomes for everyone as if they had been in the control arm, and abstinence outcomes for everyone as if they had been in the intervention arm. For person i , let P_{Ci} and P_{Ti} denote the estimated probability of abstaining in the control and intervention arms, respectively. We obtained P_{Ci} and P_{Ti} for everyone.

Among those in the principal stratum that always abstains, let E_1 and E_2 denote estimated mean withdrawal symptom scores under intervention and control conditions, respectively. Among those who never abstain, let E_3 and E_4 denote estimated mean withdrawal symptom scores under intervention and control. The treatment effect estimates become:

1. Estimated effect in those who always abstain = $E_1 - E_2$ where

$$E_1 = \frac{\sum_{i=1}^n S_{TAi} P_{Ti}}{\sum_{i=1}^n P_{Ti}} \quad E_2 = \frac{\sum_{i=1}^n S_{CAi} P_{Ci}}{\sum_{i=1}^n P_{Ci}}$$

$$u + (1-u) \frac{\sum_{i=1}^n P_{Ci}}{\sum_{i=1}^n P_{Ti}}$$

2. Estimated effect in those who never abstain = $E_3 - E_4$, where

$$E_3 = \frac{\sum_{i=1}^n S_{TNI} (1 - P_{Ti})}{\sum_{i=1}^n (1 - P_{Ti})} \quad E_4 = \frac{\sum_{i=1}^n S_{CNI} (1 - P_{Ci})}{\sum_{i=1}^n (1 - P_{Ci})}$$

$$\frac{\frac{1}{v} + (1 - \frac{1}{v}) \frac{\sum_{i=1}^n (1 - P_{Ti})}{\sum_{i=1}^n (1 - P_{Ci})}}{\sum_{i=1}^n (1 - P_{Ci})}$$

3. Estimated effect in those who abstain with the intervention only = $u \times E_1 - \frac{1}{v} \times E_4$

In this paper, we used the robust sandwich standard errors of Huber.[7] [33] However, researchers can also use bootstrap standard errors.

Other Analytic Details

In addition to estimating principal stratum causal effects, we also applied more common methods. Fisher's exact tests and T-tests were used to compare demographics between the intervention arms of the study. We used exact McNemar's tests and paired T-tests to compare changes in CO-confirmed abstinence and withdrawal symptoms between baseline and follow-up. We used simple linear and multiple linear regressions to examine the overall intervention effect on withdrawal symptoms (intention-to-treat) and the intervention effect within the observed abstainers. The variables entered into all regression models included education (entered as a categorical variable), race/ethnicity (non-Hispanic white, non-Hispanic black, other), age, baseline daily number of cigarettes smoked, baseline stated intention to quit smoking (response to question, "are you seriously thinking of quitting smoking within the next 30 days"), and baseline withdrawal symptoms.

For ease of presentation in this tutorial paper, we performed a complete case analysis in which missing data were excluded. Methods such as multiple imputation [34] can be used to account for missing data by researchers implementing principal stratification approaches. P-values less than 0.05 associated with two-sided hypothesis tests were used as the criteria of statistical significance.

Results

As described in Cropsey and colleagues,[20] 71 individuals started the intervention immediately and 289 provided waitlist control data. Of these, 33 were excluded from this study due to missing baseline data and 66 were excluded due to missing follow-up data. This left us with 58 individuals in the intervention group and 203 individuals in the control group. The control group was younger than the intervention group ($p=0.013$ by unpaired T-test), but the magnitude of the difference was not great (3 years). There were no other statistically significant differences between the two groups in baseline characteristics.

In the last two rows of table 3, we present the follow-up characteristics of the sample. There was a statistically significant change in the rate of CO-confirmed abstinence comparing baseline to follow-up in the intervention group (3% to 47%, $p<0.001$ by McNemar's test) but not when comparing baseline to follow-up in the control group (6% to 6%, $p=1.00$ by McNemar's test). Although the abstinence rate did not change in the control group (12 participants having abstinence at each time period), just five participants had evidence of abstinence at both periods. In this table, statistically significant differences in withdrawal symptoms between the two groups emerge. Withdrawal symptoms showed statistically significant increases from baseline ($p<0.003$ by paired T-tests within both groups). However, the increase was higher in the control than the intervention group ($p<0.001$ comparing the change scores between the two groups using an unpaired T-test).

We can use table 3 to estimate the proportion in each principal stratum. As shown in table 2, all of those who abstain while in the control condition are in the principal stratum that always abstains. Hence, in table 3, there are 6% in the stratum that always abstains. Also, as shown in table 2, all of those who do not abstain while in the intervention condition never abstain. Thus, there are 53% who never abstain. This leaves 41% in the principal stratum that only abstains in the intervention arm.

In table 4, we present the estimated effect of the intervention on withdrawal symptoms in different groups. In table 4a, we present the effect of the intervention on the entire sample. The intervention had a statistically significant reduction in withdrawal symptoms in the entire sample as shown in table 4a ($p<0.005$) without and with regression adjustment. The

magnitude of the effect in the entire sample was large relative to the standard deviation of the baseline withdrawal scores (greater than 0.5 standard deviations).

In table 4b, we present estimates of the effect in the subgroup observed to abstain in the sample, as per the recommended subgroup analysis of Shiffman and colleagues.[22] The effect of the intervention was not statistically significant either without or with adjustment for baseline covariates. The magnitude of the adjusted estimate in those observed to abstain in the study was very small (a reduction in withdrawal symptoms of -0.5 units). However, as depicted in tables 1 and 2, estimating an intervention effect in this subgroup is not meaningful since it represents the effect in a population that is a mixture of principal strata.

In figures 2 and 3, we also present estimates of the effect in groups categorized by potential abstinence status at plausible values of the sensitivity parameters u and v . The estimate of the intervention effect is only a function of the sensitivity parameter u in those who abstain regardless of intervention assignment. Among those who would abstain from smoking regardless of intervention assignment, the intervention reduces symptoms by an estimated -5.5 to -11.8 on the Minnesota Withdrawal Scale over a range of the sensitivity parameter u from 1 to 3; the absolute magnitude of the estimated effects are monotone increasing over this range. The results are statistically significant ($p < 0.015$) over the range of u presented. When u is equal to 1, we assume that the average withdrawal symptoms in the intervention arm of the group that can abstain only when assigned to the intervention are equal to those of the group that can always abstain. When u is equal to 3, we assume that the average withdrawal symptoms when assigned to the intervention are much lower for those who can always abstain. Since the baseline standard deviation of the Minnesota Withdrawal Scores is approximately 5.0 with a range of 0 to 32, an estimated effect of at least -5.5 (greater than 1.0 standard deviation) is clinically relevant.

The effect estimate in the group that would never have CO-confirmed abstinence from smoking regardless of intervention assignment varies over assumptions about the sensitivity parameter v . The point estimates for the effect of this intervention range from -0.8 to -7.4 over the range of the sensitivity parameter. The absolute magnitude of the point estimates are not as great as those in the group that always abstains regardless of intervention assignment. Here, the inferences are dependent on the choice of the sensitivity parameter. The estimates are only statistically significant ($p < 0.05$) when v is greater than 1.4. When v is equal to 1, we assume that the average withdrawal symptoms in the control arm of the group that can abstain only when assigned to the intervention are equal to those that can never abstain. When v is equal to 3, we assume that the average withdrawal symptoms when assigned to the control are higher for those who can never abstain.

The estimates of the intervention effect in those who would only abstain in the intervention arm are a function of both sensitivity parameters u and v . The effect estimates in this group range from -6.3 to 3.2 . The estimates are statistically significant except for a band when v equals approximately 2. Further, as v increases above 2, the direction of the effect flips and the intervention worsens symptoms.

We examined pairwise comparisons of effect estimates among principal strata for evidence that potential abstinence group membership moderated effects. The p -values for tests of moderation were dependent on the sensitivity parameters and ranged from <0.001 to 0.99. As an example of how the p -values vary over the sensitivity analysis, the p -value comparing the effect in those who always abstain versus those who abstain only when assigned to the intervention was 0.684 when both sensitivity parameters were equal to 1. However, the p -value was 0.035 when both sensitivity parameters were equal to 1.33.

We examined compliance as calculated by the percentage of used patches returned among those in the intervention group during the week of follow-up (week 4 or week 5). Forty-three of the 58 intervention participants had compliance data. Overall, there was 87% compliance, with 93% compliance among those with CO-confirmed abstinence and 83% compliance among those without abstinence ($p=0.234$ by a T-test).

In comparison, Zhang and Rubin's[5] estimates that do not make assumptions such as ours give bounds on the treatment effect among those who always abstain of -26 to -2.2 .

Discussion

In our study, traditional methods of assessing the effect of smoking cessation programs among observed abstainers failed to find a statistically significant effect of the intervention on nicotine withdrawal symptoms. By examining the effect of the intervention on withdrawal symptoms using our causal framework, we did find statistically significant effects, particularly in the group that would abstain regardless of whether they were assigned to the intervention or the control condition. Of note is that the group in which the effect was found to be most statistically significant (the always abstainers) is the group most similar to the observed abstainer groups in which Shiffman and colleagues[22] proposed evaluating symptoms.

Our hypothetical example can give insight into why our method was able to detect effects but Shiffman and colleague's[22] was not. Observed abstainers are a mixture of those who would always abstain regardless of intervention assignment and those who would only abstain if assigned to the intervention. In the figures, we see that the effect in the group that abstains only when assigned to the intervention (but does not abstain when assigned to the control) is generally less than the effect in the group that abstains regardless of intervention assignment. Hence, by naively examining the effect of the intervention on withdrawal symptoms in the observed abstainers, we could be diluting the estimate of effectiveness by mixing a group for whom the intervention might not be as effective in reducing symptoms (those who can abstain only when assigned to the intervention) with a group for whom the intervention appears to be highly effective regardless of the sensitivity parameters (those who always abstain regardless of intervention assignment). The causal analysis avoids this problem.

Our assumptions such as monotonicity greatly reduce the range of effect estimates. Zhang and Rubin's[5] method that makes no such assumptions gave a much wider range on the bounds of point estimates among those who always abstain.

We demonstrated that a behavioral intervention combined with nicotine replacement therapy can reduce withdrawal symptoms in an incarcerated population, but that inferences are dependent on assumptions about abstinence from smoking prior to reporting withdrawal symptoms. There was a clear intervention effect in reducing withdrawal symptoms among those who would abstain regardless of intervention assignment. However, the intervention effect was less pronounced and more ambiguous among those in the groups that 1) would abstain only when assigned to the intervention or 2) would never abstain regardless of intervention assignment.

Our findings concerning the difference between the principal strata make intuitive sense. For example, those who would never have CO-confirmed abstinence might not be benefiting as much from the intervention as those who would always have such abstinence, which could explain why the effect of the intervention on symptoms seems attenuated in this group. It is possible that having factors that make one less responsive to nicotine replacement therapy might reduce the ability of the intervention to reduce withdrawal symptoms in this group.

There are limitations to our substantive conclusions. One is that the sample consisted of a female prisoner population. The results of this study might not be generalizable to other groups. Also, the findings might have been different if the control condition had included a placebo-type intervention which participants expected would help them quit. Such a control intervention could have changed expectations about likely withdrawal symptoms and hence changed the way participants reported symptoms.

A focus of this paper has been to present a framework for investigating the effectiveness of a smoking cessation intervention on withdrawal symptoms. The potential outcomes framework of investigating intervention effects can help shed light on paradoxes in the relationship between withdrawal symptoms and smoking that have been pointed out in a review by Piasecki.[35] Piasecki argued that while there has not been strong evidence linking the time course of withdrawal symptoms to relapse risk, the ability of nicotine replacement therapy to reduce both withdrawal symptoms and relapse risk in a number of studies suggests that symptoms and relapse are related. In this paper, we demonstrate that the effect of smoking cessation programs on withdrawal symptoms might depend on potential CO-confirmed abstinence status. This work could be extended by examining intervention effects on cessation within groups defined by potential withdrawal symptom outcomes.

In this paper, we presented a sensitivity analysis to allow us to overcome nonidentifiable assumptions. Some researchers have advocated using stronger assumptions rather than sensitivity analyses to identify causal effects.[36][37] Still, using sensitivity analyses to estimate principal stratification effects is common.[9][14][15] Many have advocated sensitivity analyses for non-identifiable model assumptions.[38] In contrast, Pearl has proposed inequality tests to investigate violations of assumptions under his discussion of instrumental variables.[2]

While we believe that our monotonicity assumption is scientifically justifiable in this context, Shepherd and colleagues[14] demonstrated use of a stochastic monotonicity assumption that could be used instead. A limitation is that the method requires additional sensitivity parameters.

In conclusion, we have shown that a behavioral-based intervention combined with nicotine replacement therapy can reduce withdrawal symptoms, but the reduction is likely related to potential CO-confirmed abstinence prior to assessment of the withdrawal symptoms. Principal stratification allowed us to identify effects that traditional analyses would overlook. The potential outcomes framework of investigating intervention effects presented in this paper can be useful for investigating complex pathway relationships when post-randomization covariates can affect outcomes related to an intervention.

Acknowledgments

Research was supported in part by NIH grant P30 CA 06927 and an appropriation from the Commonwealth of Pennsylvania. This study was funded by a K23 award to K. Cropsey by the National Institute on Drug Abuse of the National Institutes of Health (Grant K23DA15774). Product support was provided by GlaxoSmithKline.

Appendix

Here we elucidate identification of effects. Let Z indicate treatment assignment ($Z=1$ if assigned to treatment, 0 otherwise). Let S_0 and R_0 represent an individual's withdrawal symptoms and wait-list time abstinence status when assigned to the control condition and S_1 and R_1 be the withdrawal symptoms and abstinence status when assigned to the intervention. R_0 and R_1 are both binary (1 if abstains, 0 otherwise). S_0 and S_1 are continuous. In our

study, we only observe the response in the initial assigned conditions. Thus, each individual has information on either S_0 and R_0 , or S_1 and R_1 , but not both sets. Hence, let S and R represent the observed data ($S=S_0$ and $R=R_0$ if assigned to control, and $S=S_1$ and $R=R_1$ if assigned to intervention). Let X represent the set of confounding covariates such that the set of potential outcomes (S_0, S_1, R_0, R_1) are independent of treatment assignment, Z , given X .

Our sensitivity parameters are hence,

$$u = \frac{E[S_1 | R_0=0, R_1=1]}{E[S_1 | R_0=1, R_1=1]}$$

$$v = \frac{E[S_0 | R_0=0, R_1=0]}{E[S_0 | R_0=0, R_1=1]}$$

where $E[\cdot]$ represents the expectation.

As detailed more extensively by Egleston and colleagues,[7] under the assumptions specified in the methods section, we have,

$$E[S_0 | R_0=0, R_1=1] = \frac{E[E[S | R=1, Z=0, X]P[R=1 | Z=0, X]]}{E[P[R=1 | Z=0, X]]}$$

$$E[S_1 | R_0=1, R_1=1] = \frac{\frac{E[E[S | R=1, Z=1, X]P[R=1 | Z=1, X]]}{E[P[R=1 | Z=1, X]]}}{u + (1-u) \frac{E[P[R=1 | Z=0, X]]}{E[P[R=1 | Z=1, X]]}}$$

$$E[S_1 | R_0=0, R_1=1] = \frac{E[E[S | R=0, Z=1, X]P[R=0 | Z=1, X]]}{E[P[R=0 | Z=1, X]]}$$

$$E[S_0 | R_0=0, R_1=0] = \frac{\frac{E[E[S | R=0, Z=0, X]P[R=0 | Z=0, X]]}{E[P[R=0 | Z=0, X]]}}{\frac{1}{v} + \left(1 - \frac{1}{v}\right) \frac{E[P[R=0 | Z=1, X]]}{E[P[R=0 | Z=0, X]]}}$$

The equalities show how the causal estimands are related to the observed data. We can fit a multiple linear regression to estimate conditional average withdrawal scores, $E[S | R, Z, X]$, and a logistic model to estimate conditional probabilities of abstention, $P[R=1 | Z, X]$. We then combine these estimates as described in the methods section.

Glossary of Commonly Used Terms

Identifiable

Heuristically, being able to examine the data and infer to which principal stratum an individual belongs. Since we do not observe

	individuals' outcomes under both randomization arms, we do not know to which subgroups individuals belong. Our assumptions allow us to probabilistically identify the principal strata.
Intermediate variable	A variable that lies in the causal pathway between two other variables as in figure 1. Also called a mediator variable.
Monotonicity	The assumption that no one would abstain from smoking if assigned to the control group but would not abstain if assigned to the intervention group.
Potential outcome	The outcome (either intermediate or distal) that an individual would have under a randomization arm. We assume that individuals' potential outcomes exist for all arms, but we only observe the outcome under the assigned randomization arm.
Principal Stratification / Principal Strata	In a two arm trial, the grouping of individuals by the potential outcomes that they would have under both arms of a trial. These groupings cannot be identified using the observed data.
Sensitivity analysis	Estimating effects over a range of assumptions that cannot be verified using the data at hand.

References

1. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002; 58:21–29. [PubMed: 11890317]
2. Pearl, J. *Causality: Models, Reasoning, and Inference*. Second Edition. Cambridge, UK: Cambridge University Press; 2009.
3. Lin JY, Ten Have TR, Elliott MR. Nested Markov compliance class model in the presence of time-varying noncompliance. *Biometrics*. 2009; 65(2):505–513. [PubMed: 18759831]
4. Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. *Biometrics*. 2009; 65(2): 530–538. [PubMed: 18759836]
5. Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics*. 2003; 28(4):353–368.
6. Hayden D, Pauler DK, Schoenfeld D. An estimator for treatment comparisons among survivors in randomized trials. *Biometrics*. 2005; 61(1):305–310. [PubMed: 15737107]
7. Egleston BL, Scharfstein DO, Freeman EE, West SK. Causal inference for non-mortality outcomes in the presence of death. *Biostatistics*. 2007; 8(3):526–545. [PubMed: 16980696]
8. Egleston BL, Scharfstein DO, MacKenzie E. On estimation of the survivor average causal effect in observational studies when important confounders are missing due to death. *Biometrics*. 2009; 65(2):497–504. [PubMed: 18759833]
9. Gilbert PB, Bosch RJ, Hudgens MG. Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics*. 2003; 59(3):531–541. [PubMed: 14601754]
10. Roy J, Hogan JW, Marcus BH. Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics*. 2008; 9(2):277–289. [PubMed: 17681993]
11. Gallop R, Small DS, Lin JY, Elliott MR, Joffe M, Ten Have TR. Mediation analysis with principal stratification. *Statistics in Medicine*. 2009; 28(7):1108–1130. [PubMed: 19184975]
12. Jo B. Causal inference in randomized experiments with mediational processes. *Psychological Methods*. 2008; 13(4):314–336. [PubMed: 19071997]
13. Rubin DB. Causal inference through potential outcomes an principal stratification: Application to studies with "censoring"s due to death. *Statistical Science*. 2006; 21(3):299–309.

14. Shepherd BE, Redman MW, Ankerst DP. Does finasteride affect the severity of prostate cancer? A causal sensitivity analysis. *Journal of the American Statistical Association*. 2008; 103(484):1392–1404. [PubMed: 20526381]
15. Sjolander A, Humphreys K, Vansteelandt S, Bellocco R, Palmgren J. Sensitivity analysis for principal stratum direct effects, with an application to a study of physical activity and coronary heart disease. *Biometrics*. 2009; 65(2):514–520. [PubMed: 18759834]
16. Neyman J. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*. 1990; 5:465–472. D.M., Dabrowska; T.P., Speed, editors. from the original in *Roczniki Nauk Rolniczych Tom X*. 1923. p. 1-51. As translated and edited by
17. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*. 1974; 66:688–701.
18. Holland PW. Statistics and causal inference. *Journal of the American Statistical Association*. 1986; 81:945–960.
19. Rubin DB. Which Ifs have causal answers. Commentary on Holland PW. *Statistics and Causal Inference*. *Journal of the American Statistical Association*. 1986; 81(396):961–962.
20. Cropsey K, Eldridge G, Weaver M, Villalobos G, Stitzer M, Best A. Smoking cessation intervention for female prisoners: Addressing an urgent public health need. *American Journal of Public Health*. 2008; 98(10):1894–1901. [PubMed: 18703440]
21. Stead LF, Perera R, Bullen C, Mant D, Lancaster T. Nicotine replacement therapy for smoking cessation. *Cochrane Database of Systematic Reviews*. 2008; (Issue 1) Art. No.: CD000146. DOI: 10.1002/14651858.CD000146.pub3.
22. Shiffman S, West RJ, Gilbert DG. Recommendation for the assessment of tobacco craving and withdrawal in smoking cessation trials. *Nicotine and Tobacco Research*. 2004; 6(4):599–614. [PubMed: 15370156]
23. Piasecki TM, Fiore MC, Baker TB. Profiles in discouragement: Two studies of variability in the time course of smoking withdrawal symptoms. *Journal of Abnormal Psychology*. 1998; 107(2): 238–251. [PubMed: 9604553]
24. Piasecki TM, Niaura R, Shadel WG, Abrams D, Goldstein M, Fiore MC, Baker TB. Smoking withdrawal dynamics in unaided quitters. *Journal of Abnormal Psychology*. 2000; 109(1):74–86. [PubMed: 10740938]
25. Jorenby DE, Hatsukami DK, Smith SS, Fiore MC, Allen S, Jensen J, Baker TB. Characterization of tobacco withdrawal symptoms: Transdermal nicotine reduces hunger and weight gain. *Psychopharmacology (Berl)*. 1996; 128(2):130–138. [PubMed: 8956374]
26. Piasecki TM, Jorenby DE, Smith SS, Fiore MC, Baker TB. Smoking withdrawal dynamics: II. Improved tests of withdrawal-relapse relations. *Journal of Abnormal Psychology*. 2003; 112(1): 14–27. [PubMed: 12653410]
27. Hall SM, Munoz R, Reus VI. Cognitive-behavioral intervention increases abstinence rates for depressive-history smokers. *Journal of Consulting and Clinical Psychology*. 1994; 62:141–146. [PubMed: 8034816]
28. Cropsey KL, Eldridge GD, Weaver MF, Villalobos GC, Stitzer ML. Expired carbon monoxide levels in self-reported smokers and non-smokers in prison. *Nicotine & Tobacco Research*. 2006; 8:653–659. [PubMed: 17008192]
29. Hughes JR, Hatsukami D. Signs and symptoms of tobacco withdrawal. *Archives of General Psychiatry*. 1986; 43:289–294. [PubMed: 3954551]
30. Bellamy SL, Lin JY, Ten Have TR. An introduction to causal modeling in clinical trials. *Clinical Trials*. 2007; 4(1):58–73. [PubMed: 17327246]
31. Zhang, J.; Rubin, DB.; Mealli, F. Evaluating the effects of job training programs on wages through principal stratification. In: Millimet, D.; Smith, J.; Vytlacil, E., editors. *Advances in Econometrics: Modeling and Evaluating Treatment Effects in Econometrics*. Elsevier Science Ltd, UK; 2007.
32. Manski CF, Sandefur GD, McLanahan S, Powers D. Alternative estimates of the effect of family structure during adolescence on high school graduation. *Journal of the American Statistical Association*. 1992; 87:25–37.
33. Huber PJ. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*. 1964; 35(1):73–101.

34. Schafer, JL. Analysis of incomplete multivariate data. Boca Raton, FL: Chapman & Hall/CRC; 1997.
35. Piasecki TM. Relapse to smoking. *Clinical Psychology Review*. 2006; 26:196–215. [PubMed: 16352382]
36. Joffe MM, Small D, Hsu CY. Defining and estimating intervention effects for groups that will develop an auxiliary outcome. *Statistical Science*. 2007; 22(1):74–97.
37. Elliott MR, Joffe MM, Chen Z. A potential outcomes approach to developmental toxicity analyses. *Biometrics*. 2006; 62(2):352–360. [PubMed: 16918899]
38. Scharfstein DO, Halloran ME, Chu H, Daniels MJ. On estimation of vaccine efficacy using validation samples with selection bias. *Biostatistics*. 2006; 7(4):615–629. [PubMed: 16556610]

Figure 1.
Historical depiction of intermediate variable (mediation) pathway.

Figure 2.

Intervention effects in those who always and never abstain under various assumptions about the sensitivity parameters. The estimate of the effect is statistically significant in the region in which the 95% confidence interval does not cross 0.

Legend:

$$u = \frac{\text{Average intervention withdrawal score if in group that abstains with intervention only}}{\text{Average intervention withdrawal score if in group that always abstains}}$$

$$v = \frac{\text{Average control withdrawal score if in group that never abstains}}{\text{Average control withdrawal score if in group that abstains with intervention only}}$$

Figure 3.

Intervention effect in those who abstain with the intervention only under various assumptions about the sensitivity parameters. The contour lines represent the point estimates and the shading represents the p-value.

Legend:

$$u = \frac{\text{Average intervention withdrawal score if in group that abstains with intervention only}}{\text{Average intervention withdrawal score if in group that always abstains}}$$

$$v = \frac{\text{Average control withdrawal score if in group that never abstains}}{\text{Average control withdrawal score if in group that abstains with intervention only}}$$

Table 1

Withdrawal symptom rates in a hypothetical sample by principal stratum and treatment assignment. Assumes use of dichotomous withdrawal symptom variable.

Principal Stratum	Treatment assignment	Has withdrawal symptoms (% of total)	Has no withdrawal symptoms (% of total)	Row Total %
(1) Always abstains	Control	3%	7%	10%
(1) Always abstains	Intervention	2%	8%	10%
(2) Abstains with intervention only	Control	3%	12%	15%
(2) Abstains with intervention only	Intervention	6%	9%	15%
(3) Never abstains	Control	5%	20%	25%
(3) Never Abstains	Intervention	5%	20%	25%
(4) Abstains with control only	Control	0%	0%	0%
(4) Abstains with control only	Intervention	0%	0%	0%
Column Total %		24%	76%	100%

Table 2

Observed withdrawal symptom rates from table 1 when principal stratum is not known.

Observed Abstinence Status	Treatment assignment	Principal stratum from table 1	Has withdrawal symptoms (% of total)	Has no withdrawal symptoms (% of total)	Row Total %
Observed to abstain	Control	(1) and (4)	3% + 0% = 3%	7% + 0% = 7%	10%
Observed to abstain	Intervention	(1) and (2)	2% + 6% = 8%	8% + 9% = 17%	25%
Not observed to abstain	Control	(2) and (3)	3% + 5% = 8%	12% + 20% = 32%	40%
Not observed to abstain	Intervention	(3) and (4)	5% + 0% = 5%	20% + 0% = 20%	25%
Column total %			24%	76%	100%

Table 3

Characteristics of data sample. Higher withdrawal symptom scores indicate more symptoms.

	Control	Intervention	P-value
Sample Size (with withdrawal/abstinence data)	203	58	
Baseline			
Mean baseline withdrawal symptoms (SD)	9.2 (5.3)	9.9 (5.4)	0.363
CO-confirmed abstinence at baseline	6%	3%	0.743
Mean age (SD)	32.6 (8.6)	35.7 (8.2)	0.013
Race			0.479
Non-Hispanic African American/Black	45%	45%	
Non-Hispanic White	48%	43%	
Other	7%	12%	
Education			0.084
Less than high school degree	31%	17%	
High school graduate	42%	45%	
Post-high school education	27%	38%	
Considering quitting smoking within 30 days	49%	50%	0.883
Mean baseline cigarettes smoked daily (SD)	15.7 (9.0)	16.7 (8.8)	0.439
Follow-up			
Mean withdrawal symptoms at follow-up (SD)	16.0 (7.1)	12.9 (8.2)	0.005
CO-confirmed abstinence at follow-up	6%	47%	<0.001

Table 4

Estimated effect of intervention on withdrawal symptoms in select subgroups.

Method*	Subgroup	Point Estimate	95% CI	P-value
Table 4a. Overall				
T-test	Entire sample	-3.1	-5.2, -0.9	0.005
Multiple linear regression	Entire sample	-3.6	-5.5, -1.7	<0.001
Table 4b. Observed abstainers as per Shiffman et al.[22]				
T-test	CO $\leq$ 2 ppm in sample	-4.9	-10.1, 0.2	0.059
Multiple linear regression	CO $\leq$ 2 ppm in sample	-0.5	-5.1, 4.0	0.811

* The multiple linear regressions adjust for education, race, age, baseline number of cigarettes, baseline intention to quit, and baseline withdrawal symptoms