[Presented at 12th Midwest International Economic Development Conference]

Direct Questioning or List-based Questioning:

Evidence from a Survey Experiment on Intravenous Infusion Use and Smoking in China

Yanfang Su

Harvard T. H. Chan School of Public Health

Abstract

Background Measuring health through surveys is challenging because participants may respond in a socially favorable but untruthful way. To overcome this social desirability bias, attempts have been made to measure human behaviors through complex indirect questioning methods, such as the list experiment. This study compared a list experiment questioning strategy to the standard direct questioning method for two behaviors, intravenous infusion use and smoking. It was expected that intravenous infusion use would be perceived as being socially desirable or neutral and smoking as being socially undesirable by students. The hypothesis was that indirect questioning would increase the reporting of smoking compared to direct questioning, and that the gap between indirect questioning and direct questioning would be significantly larger for smoking than for intravenous infusion use.

Methods A survey experiment was designed to measure the prevalence of intravenous infusion use and smoking among medical students in China by both direct and list-based questions. In a two-by-two design, two groups were asked to respond to a list-based control question, followed by direct questions on either smoking or intravenous infusion use. The second two groups responded to list-based questions about smoking or intravenous infusion use, followed by a direct placebo question.

Results Data were collected from 1,439 medical students. The estimated prevalence of smoking from indirect and direct questions was 4% and 8%, respectively, but the 4% negative difference was non-significant. The estimated prevalence of intravenous infusion use from indirect and direct questions was 43% and 52%, respectively, but the 9% negative difference was non-significant. The difference in differences was 5%, which was not significantly different from zero.

Conclusions The list experiment yielded lower point estimates of prevalence than direct questioning for smoking as well as intravenous infusion use, but the findings were non-significant. These findings contradict the assumption that smoking should show higher estimates using an indirect question compared to a direct question if smoking was socially undesirable. List experiments might introduce downward biases rather than alleviate them due to cognitive difficulty in responding. List experiments might not be more suitable than the anonymous self-administered direct method for measuring health behaviors.

Key words: self-reports; social desirability bias; list experiment; smoking; intravenous infusion use; China

I. Introduction

The tendency of respondents to answer survey questions in a manner that is viewed favorably by others suggests a social desirability bias (Sudman, Bradburn et al. 1996; King and Bruner 2000). Direct questioning might be more prone to social desirability bias than indirect questioning through a list experiment (i.e., item count technique). In list experiments, individual responses about sensitive topics are not collected; instead, a respondent only indicates the total number of statements that apply in the list. This study was designed to compare self-administered direct questioning and list-based questioning when measuring the prevalence of intravenous infusion use and smoking. The intent of the study was to examine the extent to which list experiments elicit distinctive prevalence levels of two health behaviors, which hypothetically have differing degrees of social desirability.

A. Social desirability

In the psychology literature, social desirability was first interpreted as a personality characteristic and the measurement of social desirability has evolved over time. Crowne and Marlowe (1960) developed a test to measure social desirability as a personality trait. The Marlowe-Crowne Social Desirability Scale consisted of 33 true/false items and generated a score indicating a high or low tendency of a person to provide socially desirable responses (Crowne and Marlowe 1960). Then, in 1991, Paulhus developed another method, the Balanced Inventory of Desirable

Responding, a questionnaire designed to measure two forms of socially desirable responses (Paulhus 1988). This 40-item instrument provided separate subscales for “impression management,” when there was a tendency not to be honest, and “self-deceptive enhancement,” when there was a tendency to give honest but inflated descriptions (Paulhus 1988). In self-evaluation on social desirability scales in China, it has been shown that for college students, the need to enhance one’s image might take precedence over the need to be honest (Liu, Xiao et al. 2003). Rather than reflecting a constant personality trait, social desirability varies by the nature of the topic.

B. The list experiment

In efforts to address social desirability bias, complex indirect survey techniques have been developed (Raghavarao and Federer 1979; Nederhof 1985; Fisher 1993). One of the most popular indirect survey methods is known as the item count technique (Droitcour, Caspar et al. 1991; Dalton, Wimbush et al. 1994) or the list experiment (Kuklinski, Cobb et al. 1997). In the list-based question, respondents indicate the total number of statements that apply to him or her in the list. It has been argued that an aggregated response to a list of statements is less sensitive than individual responses to a single question. When a respondent is asked how many statements in a list apply to them, he or she is more likely to reveal an accurate answer, even if the list contains sensitive statements. Conducted properly, the list experiment may be a more suitable tool than direct questioning when measuring sensitive health behaviors.

The design of a list experiment involves multiple parts: a key statement (i.e., the statement mentioning sensitive behavior), several non-key statements, and a placebo statement (Please refer to Table 8 in Appendix 1 for an example of a list experiment). In a treated list, a key statement is accompanied by several non-key statements. A control list is identical to the treated list, except the key statement is replaced by a placebo statement (i.e., a statement that has been determined to be highly unlikely to be true among the target population). By examining the differences in responses between the randomized treated and control lists, researchers can estimate the prevalence of the sensitive behavior.

Researchers hypothesize that the indirect survey techniques reduce social desirability bias by protecting the privacy of respondents (De Jong, Pieters et al. 2010), and there is some evidence corroborating this. For example, studies have shown that list experiments can reduce over-reporting of positively perceived behaviors such as church attendance (Presser and Stinson 1998), voter turnout (Belli, Traugott et al. 1999; Burden 2000; Holbrook and Krosnick 2010; Comșa and Postelnicu 2013) and "sense of purpose" in work motivation (Antin and Shaw 2012). List experiments can reduce under-reporting of undesirable behaviors such as abortion (Jones and Forrest 1992), drug use (Falck, Siegal et al. 1992; McNagny and Parker 1992; Fendrich and Vaughn 1994; McElrath, Dunham et al. 1995), sexual risk behavior (LaBrie and Earleywine 2000), anti-gay sentiment (Coffman, Coffman et al. 2013) and "killing time" as a work motivation (Antin and Shaw 2012).

However, in other studies using list experiments, results have been mixed. For instance, some studies found that drug use was more detectable in a list experiment than in direct questioning (Falck, Siegal et al. 1992; McNagny and Parker 1992; Fendrich and Vaughn 1994; McElrath, Dunham et al. 1995), while another study found that the behavior was equally detectable by both a list experiment and direct questioning (Droitcour, Caspar et al. 1991). Given the mixed results in the research to date, more evidence is needed to address the usefulness of the list experiment (Tsuchiya, Hirai et al. 2007).

C. Intravenous infusion use and smoking in China

Intravenous infusion use* was chosen as a target behavior because of its widespread and inappropriate use in China, as described below. Smoking was chosen as the secondary target behavior of this study because, as described below, it has been perceived as socially undesirable, allowing it to be an anchor for comparative analysis for intravenous infusion use. Specifically, this study assumes negative social desirability bias in self-reporting of smoking as well as positive or indistinguishable social desirability bias in self-reporting of intravenous infusion use. To our knowledge, the social desirability biases of these two behaviors have not yet been measured in China, and this study aims to address this gap.

It is likely that intravenous infusion use is socially desirable or neutral among the young population in China. Given that intravenous infusion is mainly used for administration of

* An intravenous infusion is the infusion of liquid substances directly into a vein from a drip chamber.

antibiotics in China (Currie, Lin et al. 2011), few microbiological tests were conducted prior to antibiotic prescribing (Hu, Liu et al. 2003). Studies have shown that doctors are incentivized to offer intravenous infusions because they are more profitable than oral medicines (Sun, Jackson et al. 2009). For example, one study showed that although health workers knew about the use of oral rehydration solution for diarrhea, intravenous infusions were frequently used to treat mild dehydration (Hesketh and Zhu 1997). Besides delivery of antibiotics to combat illness, intravenous infusions have also become more common for healthy students in highly competitive academic settings.

The World Health Organization (WHO) recommends that intravenous infusion be used only for managing extreme illness and for situations in which fluids cannot be taken orally among school children because of the potential risk and harm of intravenous infusion for children. Specifically, the intravenous route is recommended only for management of severe dehydration, septic shock, delivering intravenous antibiotics, and for when oral fluids are contraindicated (such as those with perforation of the intestine or other surgical abdominal problems) (WHO 2005; WHO 2013a). In countries with high compliance to the WHO recommendations, only very poor health status or severe situations lead to intravenous infusion use. While the immediate effectiveness of intravenous infusions compared to oral medication is recognized in China, the safety concerns proclaimed by the WHO have not been widely publicized.

Self-reported smoking is likely to be subject to social desirability bias when solicited via survey. Since 1950, more than 70,000 scientific papers have isolated the causal relationship of smoking and a wide variety of ailments, constituting the largest and best documented body of literature linking any behavior to disease in humans (CDC 1994). The WHO warned about the dangers of tobacco in a major report on global tobacco control in 2011 (WHO 2011). Although the smoking prevalence rate was decreasing in China (MOH 2006; Li, Hsia et al. 2011), China is the largest tobacco consumer in the world, with 301 million current smokers within the country (Li, Hsia et al. 2011). Further, children's positive attitude towards smoking was associated with tobacco advertisements (Lam, Chung et al. 1997). The WHO has urged bans on tobacco advertising, promotion and sponsorship (WHO 2013b). Given that anti-tobacco educational campaigns have been conducted in China for more than a decade, the awareness of harms from smoking has increased (Huang, Thrasher et al. 2014). Therefore, it is expected that there will be a greater level of reporting of smoking from a list experiment than that from direct questioning.

In this study, based on the theory of social desirability bias, we investigate whether a larger difference in measured prevalence exists between direct and list-based questions for smoking than for intravenous infusion use. The underlying rationale is that participants might face a conflict between the desire to reveal the correct answer and the desire to give the socially favorable response when reporting health behaviors. Additionally, given the cost of intravenous infusion use (Zhang, Eggleston et al. 2006; Xiao, Hou et al. 2010; Zeng and Cai 2011) and smoking (MOH 2006) to the health system in China, it is important to understand the prevalence and the social desirability of these behaviors.

II. Experimental design

A survey experiment was conducted, in which both direct questions and list-based questions were designed. All participants were randomized into four groups at the individual level to test the relative social desirability bias between intravenous infusion use and smoking.

A. Hypothesis

The null hypothesis was that indirect questioning would yield an equal difference in estimated prevalence levels from direct questioning for both behaviors, smoking and intravenous infusion use. The ex-ante alternative hypothesis was that a larger positive measured difference of prevalence would exist between list-based and direct questions for smoking than for intravenous infusion use. Specifically, using a behavior assumed to have non-negative social desirability bias (i.e., intravenous infusion use) as a comparison, the ex-ante expectation was that indirect questioning would yield a significantly higher estimated prevalence than direct questioning for a behavior with a negative social desirability bias (i.e., smoking). In the case that the alternative hypothesis was accepted, it would be inferred that intravenous infusion use was socially more acceptable than smoking.

B. Recruitment, consent and survey procedures

The experiment was carried out among 1,439 students in Xi'an Jiaotong University Medical School, Shanxi Province in northwestern China, in May and June, 2014. Only adult students aged 18 years or older were recruited for this study. The recruitment of students occurred in a classroom setting.

Each student responded to a short survey that was self-administered. In the survey, the following information was collected: program (undergraduate or not), the year started the program, hometown province, rural/urban, age, gender, father's educational level and mother's educational level, intravenous infusion use, smoking, and visit of Taiwan.

C. Survey instruments

The list experiment in this study was designed according to suggested best practices, such as using in-depth interviewing (Droitcour, Caspar et al. 1991), determining the optimal number of non-key statements (Corstange 2009; Comşa and Postelnicu 2013; Glynn 2013), and testing the underlying assumptions (Holbrook, Green et al. 2003; Martinez 2003; Blair and Imai 2012).

First, because the design of a list experiment can be improved by cognitive interviewing (Droitcour, Caspar et al. 1991), two rounds of cognitive interviews were applied during the development stages of the list experiment (Appendix 1).

Second, determining the number of non-key statements is an important component in the design of a list experiment. A simulation study showed that, as the number of non-key statements increased, the standard error of the point estimate of sensitive behavior increased (Corstange

2009). Additionally, too many statements might make it too cognitively difficult to respond. However, if the total number of non-key statements increases or if the non-key statements are negatively correlated, it is less likely that the respondent affirms or denies all non-key statements (Glynn 2013), thereby making it less likely that the respondent is forced to inadvertently reveal the answer to the key statement by having all “yes” or all “no” answers. Researchers have suggested that using four or less non-key statements in a list experiment is ideal (Tsuchiya, Hirai et al. 2007; Comşa and Postelnicu 2013).

Taking these points into account, twelve statements were designed in the phase of pretesting and four statements were selected from the pool to form two pairs of negatively correlated statements (Appendix 1). The following represents the control list that was used in the study:

• I performed better in math than Chinese in Grade 12. (Non-key statement 1)

• I fell asleep during class at least once in Grade 12. (Non-key statement 2)

• *I visited Gaoxiong, a city in Southern Taiwan, in Grade 12. (Placebo statement)*

• I practiced calligraphy in Grade 12. (Non-key statement 3)

• I spent time reading novels in Grade 12. (Non-key statement 4)

With the same four non-key statements, the placebo statement in the control list was replaced with the key statement in the treated list. The following were two key statements in the treated lists.

• *I smoked at least one cigarette in Grade 12. (Key statement about Smoking)*

• *I had an intravenous infusion, commonly known as 'dripping infusion,' in Grade 12. (Key statement about intravenous infusion use)*

The survey question was, "How many of the following statements were true for you in Grade 12? (Please indicate the total number but not which ones in particular.)" The students were instructed to write down the number with the explanation that, "The answer ranges from 0 to 5. Please fill 0 if none of the statements apply to you. Please fill 5 if all statements apply to you."

Third, there are important underlying assumptions to satisfy in the list experiment (Holbrook, Green et al. 2003; Martinez 2003; Blair and Imai 2012). Violation of these assumptions might introduce bias and yield little benefit in using a list experiment in improving the measurement of behaviors. Potential biases in the list experiment are addressed in this study and these biases are important to consider in the interpretation of results.

There are four important assumptions to consider. The first assumption (Assumption I) is a balance in randomization. The pre-intervention characteristics of two randomized groups should be the same, which can be demonstrated by showing that the demographic characteristics in the treated and the control groups are not significantly different. The second assumption (Assumption II) is that the response to non-key statements is independent from the presence of the key statement. In the case that the presence of the key statement induces the student to over-report or under-report the behaviors in the non-key statements, the imputed estimate of the target behavior would be biased. The independence between the non-key statements and the key statement also ensures the efficiency of the estimate from a list experiment, by eliminating the covariance between the non-key statements and the key statement. The third assumption (Assumption III) is that there is a truthful response to the key statement in the list experiment. It is assumed that, for socially desirable or undesirable behaviors, the student responds to the key statement truthfully, even though they might under-report or over-report in direct questioning. However, when all of the non-key statements apply or do not apply to the student, the protection of privacy in the list experiment vanishes. If the student answers ‘no’ to all non-key statements, the student might over-report socially desirable behaviors (floor effect); when the student answers ‘yes’ to all non-key statements, the student might under-report the sensitive behaviors (ceiling effect). The fourth assumption (Assumption IV) is that there is no design effect. It is required that no difference in cognitive difficulty exists in responding to the treated list and the control list. If the key statement adds significant cognitive difficulty to counting up the total number of statements, this assumption would be violated.

For direct questioning, intravenous infusion use was estimated by the following question: “Did you have an intravenous infusion, commonly known as ‘dripping infusion,’ in Grade 12?” Smoking was estimated by the following question: “Did you ever smoke at least one cigarette in Grade 12?” The placebo question was, “Did you visit Gaoxiong, a city in Southern Taiwan, in Grade 12?” Those three questions all generated binary responses of ‘yes’ or ‘no.’

D. **Randomization**

All students were randomized into four groups at the individual level in a two-by-two scheme with equal probability (Table 1). The list experiment about smoking consisted of a control list and a treatment list for smoking; the same design was used for intravenous infusion use. Therefore, 25% of the students were randomly assigned to each of the four groups in Table 1. Each student first responded to the list-based question, followed by the direct question (Table 1).

Table 1. List-based and direct responses, by randomized group

Randomized groups (1)	List-based responses (List) (2)	Direct responses (Direct) (3)
$\text{Smoking}_{\text{DirectQ}}$	$\text{List}_{\text{Control1}}$	$\text{Direct}_{\text{Smoking}}$
$\text{IV}_{\text{DirectQ}}$	$\text{List}_{\text{Control2}}$	$\text{Direct}_{\text{IV}}$
$\text{Smoking}_{\text{List}}$	$\text{Lsit}_{\text{Smoking}}$	$\text{Direct}_{\text{Placebo1}}$
IV_{List}	List_{IV}	$\text{Direct}_{\text{Placebo2}}$

Note: In column (1), $\text{Smoking}_{\text{DirectQ}}$ and $\text{IV}_{\text{DirectQ}}$ represent that smoking or intravenous infusion use was asked through direct questioning, respectively; $\text{Smoking}_{\text{List}}$ and IV_{List} represent that smoking statement or the statement about intravenous infusion use was buried in the list, respectively.
The contents of list-based and direct responses vary by randomized groups, specified in column (2) and column (3), respectively. In column (2), both $\text{List}_{\text{Control1}}$ and $\text{List}_{\text{Control2}}$ are responses to the same control list, consisting of four non-key statements and a placebo statement. $\text{Lsit}_{\text{Smoking}}$ is responses to a treated list, consisting of four non-key statements and a statement about smoking. List_{IV} is responses to the other treated list, consisting of four non-key statements and a statement about intravenous infusion use.
In column (3), DirectSmoking and DirectIV are responses to the direct question about smoking and intravenous infusion use, respectively. DirectPlacebo1 and DirectPlacebo2 are responses to the same placebo question about visiting a city in Taiwan.

III. Estimation strategy

A. Outcomes and equation form of hypothesis

The experiment was designed to ultimately estimate the prevalence levels of health behaviors, the difference of prevalence levels between direct questioning and indirect questioning, and difference-in-differences (DID) (Table 2). Accordingly, Table 2 shows the outcome measures, the indicators, and the mathematical equations.

Table 2. Outcome measures, indicators and mathematical equations

Outcome measure	Indicator	Mathematical equation
Prevalence of smoking	$Prevalence_{IndirectSmoking}$	mean ($List_{Smoking}$) - mean ($List_{ControlPooled}$)+ mean ($Direct_{PlaceboPooled}$)
	$Prevalence_{Directsmoking}$	mean ($Direct_{Smoking}$)
Difference #1	$Difference_{Smoking}$	$Prevalence_{IndirectSmoking}$ - $Prevalence_{Directsmoking}$ = [mean ($List_{Smoking}$) - mean ($List_{ControlPooled}$)+ mean ($Direct_{PlaceboPooled}$)] - mean ($Direct_{Smoking}$)
Prevalence of IV use	$Prevalence_{IndirectIV}$	mean ($List_{IV}$) - mean ($List_{ControlPooled}$)+ mean ($Direct_{PlaceboPooled}$)
	$Prevalence_{DirectIV}$	mean ($Direct_{IV}$)
Difference #2	$Difference_{IV}$	$Prevalence_{IndirectIV}$ - $Prevalence_{DirectIV}$ = [mean ($List_{IV}$) - mean ($List_{ControlPooled}$)+ mean ($Direct_{PlaceboPooled}$)] - mean ($Direct_{IV}$)
Difference #3	DID	$Difference_{IV}$ - $Difference_{Smoking}$ = ($Prevalence_{IndirectSmoking}$ - $Prevalence_{Directsmoking}$) - ($Prevalence_{IndirectIV}$ - $Prevalence_{DirectIV}$) = [mean ($List_{Smoking}$) - mean ($Direct_{Smoking}$)] - [mean ($List_{IV}$) - mean ($Direct_{IV}$)]

Note: The responses to the control list in two randomized groups are pooled by taking the average of the responses. $List_{ControlPooled} = \frac{List_{Control1} + List_{Control2}}{2}$

The responses to the placebo direct question in the randomized groups are pooled by taking the average of the responses.

$Direct_{PlaceboPooled} = \frac{Direct_{Placebo1} + Direct_{Placebo2}}{2}$

The estimated prevalence of "yes" responses to the key statement can be imputed by subtracting the mean of control list responses from the mean of the treated list responses and then adding the mean of the placebo question responses, as shown in Table 2. The following is a simple example how to calculate the prevalence of intravenous infusion use from a list experiment, with a single control list and a single placebo direct question.

$Prevalence_{IndirectIV}$
$= mean\ (List_{IV}) - mean\ (List_{Control}) + mean\ (Direct_{Placebo})$
= mean (non-key statements + IV statement) - mean (non-key statements + Placebo statement) + mean $(Direct_{Placebo})$
$= mean\ (IV\ statement) - mean\ (Placebo\ statement) + mean\ (Direct_{Placebo})$
= mean (IV statement)

Similar logic can be applied to estimate other indicators that use data from the list experiment in Table 2.

Social desirability bias was measured by the discrepancy between the means of list-based and direct responses, presented as $Difference_{IV}$ and $Difference_{Smoking}$ (Table 2).

In equation form, the null hypothesis was:

$$(Prevalence_{IndirectSmoking} - Prevalence_{Directsmoking}) - (Prevalence_{IndirectIV} - Prevalence_{DirectIV}) = 0$$

The alternative hypothesis was:

$$(Prevalence_{IndirectSmoking} - Prevalence_{Directsmoking}) - (Prevalence_{IndirectIV} - Prevalence_{DirectIV}) > 0$$

Further, for smoking, it was expected that prevalence would be higher for indirect questioning than for direct questioning, and therefore, $Prevalence_{IndirectSmoking} > Prevalence_{Directsmoking}$. For intravenous infusion use, it was expected that there would be similar or less prevalence found from indirect questioning than from direct questioning, and therefore,

$Prevalence_{IndirectIV} \leq Prevalence_{DirectIV}$.

B. Estimating prevalence, difference in prevalence levels and DID

Different estimation methods have been used for list experiments (Tsuchiya 2005; Blair and Imai 2012), and because the study design was crafted to provide insight into several important measurement questions, the following methods were used to respond to specific needs in this study. Both least square estimations (LSE) and maximum likelihood estimations (MLE) were applied in data analysis of the list experiment, prevalence differences, and DID, for which the following regression specification was used.

$$Y_i = \beta_0 + \beta_1 IV_{DirectQ} + \beta_2 Smoking_{List} + \beta_3 IV_{List} + \varepsilon_i$$

Y_i indicated a dependent variable, in which i represented the individual student. The dependent variables to estimate the prevalence, the difference of prevalences, and DID are presented in Appendix 2. $IV_{DirectQ}$, $Smoking_{List}$, and IV_{List} were dummy variables for participants who were assigned to the group for which the list included the placebo statement, smoking statement, and statement about intravenous infusion use, respectively. The dummy variables, $DirectQ_{IV}$, $List_{smoking}$, and $List_{IV}$ took the value '1' if a student got that version of the survey and '0' otherwise. The coefficients, β_1, β_2 and β_3, were the discrete difference of Y_i due to the variation of each dummy variable, respectively. The reference group was $Smoking_{DirectQ}$, for whom smoking was asked in the direct question. The mean of the dependent variable for the reference group was captured by β_0.

IV. Results

A. Recruitment and demographic characteristics

Totally, 1,489 students in Xi’an Jiaotong University Medical School were defined as the study population and invited to participate in the study in May and June, 2014. Finally, 1,439 students were recruited, with a participation rate at 97%. Among the recruited students, 1,369 students responded to the survey, with a response rate at 95%. Among the 1,369 students, 5 students that self-reported an age of 17 years old (though they claimed they were 18 years old or older in consent) were excluded in analysis. The students who enrolled for pretesting and the pilot were invited for the large-scale survey as well, due to administrative difficulty in excluding them from the anonymous survey. At the end of the survey, all students were asked, “Did you participate in this survey between November, 2013 and March, 2014,” to distinguish the previous participants. Therefore, 59 students that recalled that they responded to the survey in the pretesting and pilot stages were also excluded.

Finally, 1,305 students were included in data analysis and the demographic characteristics are presented in Table 3.

Table 3. Demographic characteristics

Variable	Obs	Mean
Age	1292	20.6
% of male students	1299	38%
% from rural China	1287	44%
% of hometown in Shanxi	1295	57%
Father edu <12yrs	1302	62%
Mother edu <12yrs	1302	71%
% undergraduates	1304	96%
% freshmen	1304	30%

B. Test of assumptions

First, there was no significant difference among the randomized groups (Assumption I). Randomization was balanced in terms of age, rural residents, hometown location, parental education, sex ratio, the percentage of undergraduates and the percentage of freshmen (Table 4).

Table 4. Pre-intervention demographic characteristics, by randomized group

	$DirectQ_{smoking}$	$List_{smoking}$	$DirectQ_{IV}$	$List_{IV}$	Prob > F
Mean age	21	20	21	21	>0.05
% of students from rural	43%	45%	43%	40%	>0.05
% of hometown in Shanxi	57%	62%	58%	53%	>0.05
Father's edu < 8 yrs	60%	62%	66%	58%	>0.05
Mother's edu < 8 yrs	70%	70%	74%	68%	>0.05
% of male students	38%	41%	38%	37%	>0.05
% of undergraduate students	96%	96%	96%	96%	>0.05
% of freshmen	29%	31%	31%	29%	>0.05
Observations	344	345	338	337	

Second, regarding dependence between the key statement and the non-key statements (Assumption II), in pretesting, Pearson chi-square correlation was conducted between intravenous infusion use and reading novels; no significant correlation was found between the responses to those two behaviors. However, the non-key statements were changed after cognitive interviews; thus, the correlation between behaviors of interest and the other non-key statements remained unknown.

Third, for Assumption III (truthful responses), the null hypothesis was that the percentage of students who answered ‘0’ or ‘5’ in the treated group was greater or equal to that in the control list. The distribution of responses from ‘0’ through ‘5’ is presented in Table 5.

Table 5. The proportion of students by response value

Response value	Control list			Treated list	
	Control1	Control2	Pooled Control	Smoking in the list	IV in the list
0	2%	3%	3%	3%	2%
1	7%	9%	8%	9%	6%
2	34%	33%	33%	29%	24%
3	42%	42%	42%	42%	37%
4	13%	13%	13%	15%	23%
5	1%	1%	1%	2%	8%
Obs	344	345	689	336	337

T-tests were conducted between the pooled control list and two treated lists and we failed to reject the null of truthful responses. Specifically, in testing for the floor effect, there was no significant difference between the control and the treated groups ($p>0.05$); in testing for the ceiling effect, there was no significant difference between the control and treated list about smoking ($p>0.05$) and the percentage of responses with value 5 was significantly higher in the treated list about intravenous infusion use ($p=0.00$).

In sum, the list experiment met the standards of Assumptions I (balance in randomization) and III (truthful response to the key statement). Assumption II (independence between key statement and the non-key statements) was partially tested. Assumption IV (design effect) could not be adequately assessed in this study.

C. Main results

It was important to first examine the characteristics of list-based and placebo responses.

List-based responses - The mean of list-based responses was 2.98, 2.60, 2.62 and 2.55 for the treated list about intravenous infusion use, the treated list about smoking and two control lists, respectively. The difference in the mean of estimates was 0.07 between two randomized groups responding to the control list, but it was not statistically significant ($p > 0.1$). Thus, the responses to the control list in two randomized groups can be pooled and were pooled for the analysis of the main results.

Placebo responses - For the placebo question, "Did you visit Gaoxiong, a city in Southern Taiwan, in Grade 12," 1.5% and 2.5% of the students reported they visited Taiwan, in two randomized groups, respectively. The difference between the means of placebo responses was 0.01 but it was not statistically significant ($p > 0.1$). The placebo responses can be pooled and were pooled in estimating the main results.

The estimates from direct questioning, indirect questioning, the difference of prevalence levels between the two survey methods, and difference-in-differences are presented in Table 6. There was no missing data in direct questioning and there were only two missing values in indirect questioning, among 1,305 students.

Table 6. Prevalence, difference in two prevalence levels, and DID

		Models					
		Least Square			Maximum likelihood		
Indicator	Obs	B(%)	SE	P-value	B(%)	SE	P-value
$\text{Prevalence}_{\text{IndirectSmoking}}$[1]	1308	4%	0.07	0.60	4%	0.55	0.95
$\text{Prevalence}_{\text{Directsmoking}}$[2]	325	8%	0.01	0.00	8%	0.01	0.00
$\text{Difference}_{\text{Smoking}}$[3]	1308	-4%	0.07	0.55	-4%	0.56	0.94
$\text{Prevalence}_{\text{IndirectIV}}$[1]	1308	43%	0.07	0.00	43%	0.79	0.56
$\text{Prevalence}_{\text{DirectIV}}$[2]	331	52%	0.03	0.00	52%	0.03	0.00
$\text{Difference}_{\text{IV}}$[3]	1308	-9%	0.08	0.26	-9%	0.84	0.92
DID[4]	1310	5%	0.09	0.58	5%	0.26	0.85

Note:
[1] $\text{Prevalence}_{\text{Indirect}}$: The point estimate of prevalence from indirect questioning is calculated by subtracting the mean of control list responses from the mean of the treated list responses and adding in the mean of the response to the placebo question.
[2] $\text{Prevalence}_{\text{Direct}}$: The point estimate of prevalence in direct questioning is the mean of direct responses.
[3] Difference = $\text{Prevalence}_{\text{Indirect}}$ - $\text{Prevalence}_{\text{Direct}}$
[4] Difference-in-differences (DID) = $\text{Difference}_{\text{Smoking}}$ - $\text{Difference}_{\text{IV}}$

The point estimates were consistent between least square (LS) and maximum likelihood (ML). However, contradictory to the literature (Blair and Imai 2012; Comşa and Postelnicu 2013; Meng, Pan et al. 2014), the ML estimators yielded larger standard errors in some cases.

Estimated prevalence levels from two methods - The health behaviors were measured both from direct questioning and the list experiment. From the list experiment, the estimated smoking prevalence was 4%, using the pooled control list as the reference group, and the estimator was non-significantly different from zero. From direct questioning, estimated smoking prevalence was 8%, which was significantly different from zero (Table 6). From direct questioning, the estimated intravenous infusion use was 52%. From the list experiment, the estimated intravenous infusion use was 43%, using the pooled control list as the reference group. Both point estimates were significantly different from zero (Table 6).

Difference of prevalence levels between direct questioning and indirect questioning – The difference in prevalence levels for smoking was 4%, which was negative in sign and non-significant (Table 6); the difference in prevalence levels for intravenous infusion use was 9%, which was negative in sign and non-significant.

Difference-in-differences – There was an approximately 5% difference between the two measurements and two behaviors but the estimator was non-significantly different from zero (Table 6).

V. Discussion and limitations

A survey experiment was conducted to explore the self-reported prevalence of intravenous infusion use and smoking, with the expectation that indirect questioning would reduce underreporting of smoking. The main finding was that the difference-in-differences between direct questioning and indirect questioning for two health behaviors was 5%, which was non-significant. The results failed to reject the null hypothesis that the reporting gap between direct questioning and indirect questioning was the same for intravenous infusion use and smoking among medical students in China.

It was surprising that lower estimates were yielded in the list experiment than direct questioning for smoking. There are several sources of bias that may have influenced this result. First, there is the potential bias resulting from the violation of assumptions. It was estimated that bias was very unlikely to be introduced due to unbalanced randomization (violation of Assumption I), or untruthful responses to the key statement in the list experiment (violation of Assumption III).

The main concern was violation of design effect (Assumption IV). In this study, it was very likely that the measurement error with a downward bias, which occurred in estimating the prevalence levels for both behaviors, was due to counting difficulty. More specifically, it might be sufficiently more difficult to memorize the affirmative answers and add them up in responding to the treated list compared to the control list. Other studies showed that participants' cognitive difficulties in memorizing the affirmative answers and then adding them up introduced measurement errors (Biemer, Jordan et al. 2005; Tsuchiya, Hirai et al. 2007). Other researchers have found similar results. For instance, Droitcour et al. as well as LaBrie and Earleywine applied an unmatched list experiment and they yielded lower estimates in the list experiment than direct questioning for intravenous drug use (Droitcour, Caspar et al. 1991) and college students getting drunk (LaBrie and Earleywine 2000). It was very likely that cognitive difficulty was greater in responding to the treated list than to the control list because there was an additional statement in the unmatched list experiment.

Another concern was that the response to non-key statements was dependent on the presence of the key statement, leading to a violation of Assumption II. Because this assumption was only partially tested, it is necessary to discuss the likelihood of correlation between the key statement and the non-key statements. The statement of interest was placed in the middle of the list, as the third one out of five. It was possible that the responses to the statements followed by the statement of interest were impacted by the key statement due to order effect (McClendon and O'Brien 1988; Buckley 2008; Lee, Schwarz et al. 2014). It was tested and shown that there was no significant correlation between intravenous infusion use and reading novels in the pretesting phase. However, it was left unknown whether there is a correlation between intravenous infusion use and calligraphy practice as well as between smoking and two non-key statements (i.e., calligraphy practice and reading novels) in the list.

It was also surprising that the smoking prevalence levels estimated through both survey methods in this study were lower than the estimated prevalence in the Global Adult Tobacco Survey (GATS). The GATS sampled from 100 counties/districts in China in 2010, and the estimated prevalence was 18% [95% confidence interval (14.7, 21.6)], among those 15 to 24 years old (Li, Hsia et al. 2011). In this study, the sample was medical students in Xi'an Jiaotong University, with an average age of 20 years old, and smoking prevalence was estimated for the year of 2012. The smoking prevalence was 8% and 4% using the direct and indirect questioning methods, respectively. There are several possible explanations for this discrepancy. First, it is possible that there were fewer smokers in the medical school in this study than in the nationwide sample. Second, as smoking prevalence declines over time (MOH 2006; Li, Hsia et al. 2011), the estimated prevalence in 2012 could be lower than that in 2010. Third, the cognitive difficulty of responding to the list experiment may have placed a downward bias on the estimate. Relatively, the estimate from direct questioning is closer to the national average estimate than that from the list experiment.

In this study, the results suggest that the list experiment may not be useful in improving the measurement of intravenous infusion use and smoking. Given that there are mixed results from list experiments in the literature, the results from this study belong to the pool of research that has shown no difference between the estimates from a list experiment and direct questioning for the following behaviors: intravenous drug use (Droitcour, Caspar et al. 1991), receptive anal intercourse (Droitcour, Caspar et al. 1991), college students getting drunk (LaBrie and Earleywine 2000), past engagement in counterproductive behaviors (Ahart and Sackett 2004), the prevalence of cocaine use (Biemer, Jordan et al. 2005), giving blood (Tsuchiya, Hirai et al. 2007), and condom use (Jamison and Karlan 2011). Further, counter-intuitive results have been generated from list experiments. For instance, the number of sexual partners was reported higher in direct questioning than in list-based questioning (Jamison and Karlan 2011). In such cases, the ex-anti prior about a specific behavior or the potential bias in a list experiment needs to be examined.

In the field of survey research, perhaps the most effective approach, and the path with minimal levels of social desirability bias, is the use of anonymous, self-administered direct questioning. A survey on sensitive questions could be self-administered, web-based or telephone-based rather than interviewer-administered so as to avoid interpersonal interaction (Nederhof 1985; Johnson, Hougland et al. 1989). In prior research, when participants were asked to report socially undesirable behavior in a survey free of interviewer presence as opposed to with an interviewer in the room, socially undesirable behavior was reported more frequently when the interviewer was *not* in the room (Kaminska and Foulsham 2013). This suggests that the likelihood of underreporting a socially undesirable behavior is higher when responding to another person as opposed to when in isolation. Furthermore, a recent report examining online panels by the American Association for Public Opinion Research concluded that, regardless of design, there were higher reports of socially undesirable attitudes and behaviors in self-reported web-based questionnaires than in face-to-face interviews (Baker, Blumberg et al. 2010). In this study, both self-administration of the survey and response without an identifier protected privacy. The low percentage of item non-response suggests that privacy is protected in anonymous self-administration of surveys.

There are several important limitations to this study. First, given that the study sample was medical students, it is difficult to generalize the findings to students in the general population. Second, the prevalence levels of intravenous infusion use and smoking were only measured by surveys rather than objective measurements; therefore, the validity of the survey instruments remains unknown. Third, the surveys were self-administered by participants; therefore, the results of this study cannot be generalized to other survey modes, such as interviewer administration. Fourth, results about cognitive difficulty in responding to the list-based question may not be applicable to other list experiments with less than four non-key statements.

VI. Conclusion

List experiments might not be more suitable than the anonymous self-administered direct method for measuring health behaviors. There was no evidence that list-based questioning yielded greater reports of smoking use when compared to direct questioning. Nor was evidence generated about the level of social desirability for smoking and intravenous infusion use among medical students in China.

The results from this study contradicted the ex-ante assumption that smoking should show a higher estimate of prevalence using list-based questioning than that from direct questioning if smoking was socially undesirable. The surprising finding suggests that the list-based method might introduce downward bias. The bias was plausibly due to the violation of the "no design effect" assumption.

It needs to be acknowledged that it can be a complex task, for participants in a list experiment, to count and memorize the affirmative answers. Even though the number of statements was the same in the control and the treatment groups in the list experiment in this study, the key statement about smoking or intravenous infusion use was more likely to yield an affirmative answer than the placebo statement about visiting Taiwan. Therefore, students in the treated group might experience more counting difficulty in adding up all affirmative answers.

Competing interests

The author declares that she has no competing interests.

Acknowledgements

I am thankful to Professor Zhongliang Zhou who helped with study implementation and data entry. I am thankful to students in Xi'an Jiaotong University Medical School, China, who participated in the study. I have benefited from helpful comments by Margaret McConnell, Joshua Salomon, William Hsiao, Julian Jamison, Adam Glynn, Chase Harrison, Jesse Heitner, and Jennifer Pan, as well as participants in the Doctoral Research Seminar at Harvard and the North East Universities Development Consortium. I am indebted to the B&P Foundation (Hong Kong) for financial support.

Human subject research ethics

This study was reviewed and approved by IRB in Xi'an Jiaotong University Medical School (ID: 2013-231).

Appendix 1. Paper I: Design of survey instruments, pilot, and power calculation

Pretesting to design survey instruments (n=39), cognitive interviews (n=8), and pilot (n=54) were conducted before the survey experiment. Pretesting was conducted in Xi'an Jiaotong

University on Nov. 8th, 2013. In total, 39 freshmen were enrolled for pretesting, with two focuses: potential recall issues and the design of statements in the list experiment.

Intravenous infusion – Students were asked about intravenous infusion use in different time frames. It was shown that 23% of freshmen had limited understanding about intravenous infusion. Accordingly, it was added that “intravenous infusion is commonly known as ‘dripping infusion’.” According to Table 7, about a quarter of freshmen could not recall the utilization of intravenous infusion in elementary school. The percentage dropped to 5% for recalling in senior high school. This was consistent with the intuition that it was more challenging to recall remote events compared to more recent events. According to the estimations in Table 7, recalled intravenous infusion use declined over time, from elementary school to senior high school. It was most reasonable and feasible to ask about the use of intravenous infusion in Grade 12 because it bore the least recall difficulty.

Table 7. Intravenous infusion use among students

	% among all participants	% cannot recall	% among the recalled
Elementary school	56	23	73
Junior high school	67	8	72

Senior high school	55	5	58
Grade 12	43	10	48

Non-key statements and placebo statement -- In pretesting, 12 statements were designed as the candidates for the list experiment. These were:

1. Did you do any household work in grade 12?
2. Did you read love novels in grade 12?
3. Did you read knight novels in grade 12?
4. Did you like team sports in grade 12?
5. Were TV programs about nature your favorite in grade 12?
6. Did you prefer pop music to traditional music in grade 12?
7. Did you like calligraphy in grade 12?
8. Could you see the blackboard clearly from the last row in the classroom in grade 12?
9. Did your family have a house in Hong Kong while you were in grade 12?
10. Was math your favorite course in grade 12?
11. Did you prefer word puzzles to numeric puzzles in grade 12?
12. Were you a communist party member in grade 12?

The mean of each statement and the correlation of statements were estimated to select non-key statements and a placebo question. The statement about party membership in XJU was not a good candidate for a list experiment as only 5% of the students were party members in Grade 12. Meanwhile, the test for the statement about ownership of a house in Hong Kong showed a mean of 0 and no variation. It was a good choice as a placebo question. Pearson correlations were conducted for the variable of interest (i.e., intravenous infusion in grade 12) and the ten remaining candidate statements. Freshmen who liked pop music were more likely to use intravenous infusions. Preference of pop music was negatively associated with preference of calligraphy. The preference of math was negatively associated with the preference of word puzzles; however, the correlation was not statistically significant.

In sum, according to the feedback from pretesting, the students would be asked about the use of intravenous infusion in Grade 12, in a large-scale randomized survey experiment. Four non-key statements were chosen for the list experiment, i.e., preference of classical music, preference of calligraphy, preference of math course and preference of word puzzle. House ownership in Hong Kong was chosen as the placebo.

A. Cognitive Interviewing

Cognitive interviews were conducted with a focus on the comprehension, judgment and response to the list-based question. The first round of cognitive interviews was conducted among four students in XJU Medical School, November 24th, 2013. Concurrent cognitive debriefing was conducted without specific probes. One of the participants circled all of the statements available to her, including use of intravenous infusions, showing that she was comfortable revealing her choices. The participant said that she would prefer being asked to directly circle the specific statements on the list. She complained that the instructions suggested that she had to count all the statements that applied to her in order to answer the question about total number of statements.

Two participants indicated that the statement, ‘my family owns a house in Hong Kong,’ was surprising. One participant wondered why the researcher wanted to know this piece of information. The statement was changed to the following: “I have visited Gaoxiong, a city in southern Taiwan.” The overall survey was commented as, “too simple to be true.” One of the participants was wondering what kind of research could be done with such a simple survey. Another participant declared that this was the simplest survey he had experienced.

The second round of cognitive interviews was conducted among four students in XJU Medical School, between December 30th, 2013 and January 2nd, 2014. Retrospective cognitive debriefing was applied, in which all students first answered the questionnaire and then they were invited to think aloud about the process of surveying in a private setting. Specific probes were designed for cognitive debriefing.

Probe 1: “How did you reach the answer in the list question?”

The list questions in Table 8 were tested through thinking-aloud.

Table 8. The list questions

Version	Question	Answer

A	How many of the following statements were true for you in Grade 12? (Please indicate the total number but not which ones in particular.) • Among all courses in Grade 12, my favorite was math. • I preferred pop music to classical music in Grade 12. • I visited Gaoxiong, a city in Southern Taiwan, in Grade 12. • I liked calligraphy in Grade 12. • I preferred word puzzles to numeric puzzles in Grade 12.	0 true statement 1 true statements 2 true statements 3 true statements 4 true statements 5 true statements
B	How many of the following statements were true for you in Grade 12? (Please indicate the total number but not which ones in particular.) • Among all courses in Grade 12, my favorite was math. • I preferred pop music to classical music in Grade 12. • I had intravenous infusion, commonly known as 'dripping infusion', in Grade 12. • I liked calligraphy in Grade 12. • I preferred word puzzles to numeric puzzles in Grade 12.	0 true statement 1 true statements 2 true statements 3 true statements 4 true statements 5 true statements

Student #1 and Student #2 had no problem with the question. They specifically commented on each statement and elaborated the reason why it did or did not apply to them. Student #3 recommended to change the first statement from, "Among all courses in Grade 12, my favorite was math," to, "I prefer math course to Chinese course in Grade 12," to reduce the cognitive difficulty to complete comparison of all courses. She commented that the revised statement matched better with the fifth statement.

Student #3 commented that she had no exposure to music in Grade 12 and the statement did not apply to her. Student #4 said that only one statement applied to him and it was a straightforward question to him. Student #4 circled two answers in the list-based question. When it was pointed out that she selected "1 true statement" and "4 true statements," she explained that she misunderstood it as, "1st statement is true," and, "4th statement is true," because the answers were parallel to the statements in the Chinese version of the survey. She confessed that she did not pay any attention to the sentence, “Please indicate the total number but not which ones in particular.” She read through all the statements and circled two answers. She recommended re-formatting the answer as “____true statements” for participants to fill out.

Probe 2: “What is the purpose of this study?”

It was the first time for all students to experience a list experiment. Students had no idea about the purpose of the study. When the complementary survey questionnaire was shown and the purpose of the design was introduced, those four students commented that it was an interesting way to survey intravenous infusion use. Student #2 had no idea about the purpose of the study, and, after a second thought, he said that maybe it was about the folk exchange between mainland China and Taiwan. Student #3 thought this was about whether a student was rational or more emotional in judgment and she pretended to be rational (e.g., like math course, like numerical puzzles).

Thinking-aloud: “Would you please talk a little bit more about intravenous infusion?”

Student #1 specified that he answered “yes” to the question about intravenous infusion use because he recalled that he had a severe illness in Grade 12 and used intravenous infusion. Student #2 said that he had no specific memory of intravenous infusion in Grade 12 and his best guess was that he probably had experience with it. Student #2 commented that intravenous infusion was not sensitive for him and he would like to reveal the true answer even if he was asked about this topic directly. When the topic was substituted with sexual behavior, he said that he would prefer to skip the question. Student #3 claimed that she had no intravenous infusion in Grade 12. For Student #4, his mother was a physician and he had sufficient knowledge about intravenous infusion.

The first three main adjustments made to the list-based question were based on the feedback from cognitive interviews and the last two adjustments were based on feedback from the research committee at Harvard. Therefore, all the statements were more coherent and closer to student life.

1. The instruction, "Please indicate the total number but not which ones in particular," was bolded.
2. The non-key statement, "Among all courses in Grade 12, my favorite was math," was changed into, "I prefer math course to Chinese course in Grade 12".
3. The answer format was changed from circling a number to writing down the number. The range of the answer was specified and the examples of answer 0 and 5 were given.
4. The non-key statements were changed from preference to actual behaviors.
5. The statement, 'I fell asleep during class at least once in Grade 12,' was added to make the statements on intravenouss infusions and smoking less obvious.

B. Pilot and Power Calculation

Using the finalized version of the questionnaire, the pilot was launched and 54 students were recruited in March, 2014. The estimated prevalence of intravenous infusions was 29% and 39% from direct questioning and indirect questioning, respectively. The estimated prevalence of smoking was 7% and 43% from direct questioning and indirect questioning, respectively. The DID was 26%. According to the power calculation, the number of enrolled participants needed was around 1,250 adult students (Table 9), with 80% power, alpha of 0.05, and using a one-sided test. Assuming that 5% of the students in universities were under 18 years old and the participation rate was 95%, it was planned to screen around 1,385 students in order to enroll 1,250 adult students.

Table 9. Results from power calculations

Sample size	Powers: Control list in $DirectQ_{smoking}$	Powers: Control list in $DirectQ_{IV}$
500	43%	40%
600	49%	46%
1000	71%	67%
1250	80%	77%
1500	87%	84%
2000	94%	92%
2500	98%	97%

Appendix 2. Paper I: Construction of dependent variables

The mathematical procedures involved in constructing dependent variable Y_{ji} are presented in this appendix.

The regression specification is:

$$Y_{ji}=\beta_{j0} + \beta_{j1}\, IV_{DirectQ} + \beta_{j2}\, Smoking_{List} + \beta_{j3}\, IV_{List} + \varepsilon_{ji},$$

in which j indicates the different constructions of dependent variable.

The data for regression are summarized in Table 1. The list-based responses were discrete data, ranging from 0 to 5. The direct responses were binary data, and particularly, the direct responses to the placebo question were with mean close to zero by design. The mathematical equations are presented in Table 2.

A. Indirect estimates of prevalence: Smoking and intravenous infusion use

$$\begin{aligned} & Prevalence_{IndirectSmoking} \\ &= mean\ (List_{Smoking} - List_{ControlPooled} + Direct_{PlaceboPooled}) \\ &= mean\ (List_{Smoking} - \frac{List_{Control1} + List_{Control2}}{2} + \frac{Direct_{Placebo1} + Direct_{Placebo2}}{2}) \\ &= -\ mean\ (\frac{List_{Control1}}{2}) - mean\ (\frac{List_{Control2}}{2}) + mean\ (\frac{Direct_{Placebo1}}{2}) + mean\ (List_{Smoking} + \frac{Direct_{Placebo2}}{2}) \end{aligned}$$

Equation 1.1.

Y_{1i} is constructed to estimate prevalence of smoking from the list experiment in the following manner:

$$Y_{1i} = \begin{cases} \frac{List_i}{2}, & if\ \text{Smoking}_{\text{DirectQ}} = 1 \\ \frac{List_i}{2}, & if\ \text{IV}_{\text{DirectQ}} = 1 \\ List_i + \frac{Direct_i}{2}, & if\ \text{Smoking}_{\text{List}} = 1 \\ \frac{Direct_i}{2}, & if\ \text{IV}_{\text{List}} = 1 \end{cases}$$

In which, $List_i$ is the response from a list-based question and $Direct_i$ is the response from a direct question presented in Table 1.

Therefore, Equation 1.1, with re-arrangement, becomes

$\text{Prevalence}_{\text{IndirectSmoking}}$
$= - (Y_{1i} \mid \text{Smoking}_{\text{DirectQ}}=1) - (Y_{1i} \mid \text{IV}_{\text{DirectQ}}=1) + (Y_{1i} \mid \text{Smoking}_{\text{List}}=1) + (Y_{1i} \mid \text{IV}_{\text{List}}=1)$
$= - \beta_{10} - (\beta_{10} + \beta_{11}) + (\beta_{10} + \beta_{12}) + (\beta_{10} + \beta_{13})$
$= - \beta_{11} + \beta_{12} + \beta_{13}$

Similarly, Y_{2i} is constructed to estimate prevalence of intravenous infusion use from the list experiment in the following manner:

$$Y_{2i} = \begin{cases} \frac{List_i}{2}, & if\ \text{Smoking}_{\text{DirectQ}} = 1 \\ \frac{List_i}{2}, & if\ \text{IV}_{\text{DirectQ}} = 1 \\ \frac{Direct_i}{2}, & if\ \text{Smoking}_{\text{List}} = 1 \\ List_i + \frac{Direct_i}{2}, & if\ \text{IV}_{\text{List}} = 1 \end{cases}$$

$\text{Prevalence}_{\text{IndirectIV}} = - \beta_{21} + \beta_{22} + \beta_{23}$

B. Difference of prevalence levels: Smoking and intravenous infusion use

$\text{Difference}_{\text{Smoking}}$
$= \text{mean}(\text{List}_{\text{Smoking}} - \text{List}_{\text{ControlPooled}} + \text{Direct}_{\text{PlaceboPooled}}) - \text{mean}(\text{Direct}_{\text{Smoking}})$
$= \text{mean}(\text{List}_{\text{Smoking}} - \frac{List_{Control1} + List_{Control2}}{2} + \frac{Direct_{Placebo1} + Direct_{Placebo2}}{2}) - \text{mean}(\text{Direct}_{\text{Smoking}})$

$= -\text{ mean }(\frac{List_{Control1}}{2} + Direct_{Smoking}) - \text{ mean }(\frac{List_{Control2}}{2}) + \text{ mean }(List_{Smoking} + \frac{Direct_{Placebo1}}{2}) + \text{ mean }(\frac{Direct_{Placebo2}}{2})$

Equation 1.2.

Y_{3i} is constructed to estimate the difference of prevalence levels from indirect questioning and direct questioning, for smoking, in the following manner:

$$Y_{3i} = \begin{cases} \frac{List_i}{2} + Direct_i, & if\ \text{Smoking}_{\text{DirectQ}} = 1 \\ \frac{List_i}{2}, & if\ \text{IV}_{\text{DirectQ}} = 1 \\ List_i + \frac{Direct_i}{2}, & if\ \text{Smoking}_{\text{List}} = 1 \\ \frac{Direct_i}{2}, & if\ \text{IV}_{\text{List}} = 1 \end{cases}$$

Therefore, Equation 1.2, with re-arrangement, becomes

$\text{Difference}_{\text{Smoking}}$
$= -(Y_{3i} \mid \text{Smoking}_{\text{DirectQ}}=1) - (Y_{3i} \mid \text{IV}_{\text{DirectQ}}=1) + (Y_{3i} \mid \text{Smoking}_{\text{List}}=1) + (Y_{3i} \mid \text{IV}_{\text{List}}=1)$
$= -\beta_{30} - (\beta_{30} + \beta_{31}) + (\beta_{30} + \beta_{32}) + (\beta_{30} + \beta_{33})$
$= -\beta_{31} + \beta_{32} + \beta_{33}$

Similarly, Y_{4i} is constructed to estimate the difference of prevalence levels from indirect questioning and direct questioning, for intravenous infusion use, in the following manner:

$$Y_{4i} = \begin{cases} \frac{List_i}{2}, & if\ \text{Smoking}_{\text{DirectQ}} = 1 \\ \frac{List_i}{2} + Direct_i, & if\ \text{IV}_{\text{DirectQ}} = 1 \\ \frac{Direct_i}{2}, & if\ \text{Smoking}_{\text{List}} = 1 \\ List_i + \frac{Direct_i}{2}, & if\ \text{IV}_{\text{List}} = 1 \end{cases}$$

$\text{Difference}_{\text{IV}} = -\beta_{41} + \beta_{42} + \beta_{43}$

C. Difference-in-differences

DID
= $\text{Difference}_{\text{Smoking}}$ - $\text{Difference}_{\text{IV}}$
= [mean ($\text{List}_{\text{Smoking}}$) - mean ($\text{Direct}_{\text{Smoking}}$)] - [mean ($\text{List}_{\text{IV}}$) - mean ($\text{Direct}_{\text{IV}}$)]
= - mean ($\text{Direct}_{\text{Smoking}}$) + mean ($\text{Direct}_{\text{IV}}$) + mean ($\text{List}_{\text{Smoking}}$) - mean ($\text{List}_{\text{IV}}$)
Equation 1.3.

Y_{5i} is constructed to estimating the difference-in-differences in the following manner:

$$Y_{5i} = \begin{cases} List_i, & if\ \text{Smoking}_{\text{List}} = 1\ or\ \text{IV}_{\text{List}} = 1 \\ Direct_i, & if\ \text{Smoking}_{\text{DirectQ}} = 1\ or\ \text{IV}_{\text{DirectQ}} = 1 \end{cases}$$

Therefore, Equation 1.3, with re-arrangement, becomes
DID
= - (Y_{5i} | $\text{Smoking}_{\text{DirectQ}}$=1) + ($Y_{5i}$ | $\text{IV}_{\text{DirectQ}}$=1) + ($Y_{5i}$ | $\text{Smoking}_{\text{List}}$ =1) - (Y_{5i} | IV_{List} =1)
= - β_{50} + (β_{50} + β_{51}) + (β_{50} + β_{52}) - (β_{50} + β_{53})
= β_{51} + β_{52} – β_{53}

Y_{1i},Y_{2i}, Y_{3i}, and Y_{4i} were all best fit by Gamma distributions. Therefore, the variance function used a Gamma model and the link function was Log in MLE. In estimating the difference-in-differences, Y_{5i} was under a Negative Binomial distribution. Therefore, the variance function used a Negative Binomial model and the link function was Log in MLE. For all five estimated outcomes, the p-values are generated by the “lincom” command in STATA version 12.