Draft: November 19, 2014

Accounting for Inaccuracies in Retrospective Data: A Monte Carlo Study of Smoking Cessation[1]

Donald Kenkel
Cornell University and NBER, Ithaca, NY

Joseph LeCates
Analysis Group, Boston, MA

Feng Liu
Shanghai University of Finance and Economics, Shanghai, China

Abstract

Even when contemporaneous data do not exist, retrospectively collected information provides opportunities to analyze the determinants of past behaviors. However, such data are often plagued by "heaping," the tendency of survey responses to be concentrated at certain values due to recall bias. In the context of a discrete-time hazard model, heaping results in misclassification of a binary outcome, which may substantially bias estimated regression coefficients and marginal effects. We present a model of the heaping process in a discrete-time hazard setting with significant heaping in respondent recall: smoking cessation. The 2002 Current Population Survey Tobacco Use Supplement provides a basis for a Monte Carlo analysis to quantify the bias introduced from heaping and compare several methods proposed to account for such bias. Results suggest that the bias in estimated regression coefficients and marginal effects in a discrete-time hazard model of smoking cessation are modest: less than 7% difference for estimated coefficients and 5% difference for marginal effects. Methods intending to account for such bias typically performed worse than the naïve model that ignores heaping. However, additional simulations of a policy intervention suggest that under some circumstances the biases from heaping are substantial.

Keywords

Retrospective Data, Discrete-Time Hazard Models, Smoking Cessation, Misclassification, Monte Carlo Simulation

[1] This research was supported by Award # R01 HD048828 from the National Institutes of Health. A previous version of this paper was presented at the American Society of Health Economists Third Biennial Conference and the Annual Health Econometrics Workshop. We thank John Mullahy, Colin Cameron, Tom Mroz, Mark Lewis, and session participants for their comments. We also thank our colleagues Dean Lillard and Alan Mathios who have provided useful discussions about errors in retrospective smoking data. The views presented herein are our own and do not necessarily reflect those of our home institutions; any errors are our own.

1. Introduction

Hazard models are often estimated using survey data where respondents report past behavior For example, respondents may directly or indirectly provide a start and end date for a particular behavior of interest that allows researchers to construct a synthetic longitudinal dataset with an observation for each period the respondent remains at risk (Jenkins, 1995). Because true panel data is costly to collect, reconstructed discrete-time hazard data from retrospective questions offer a low-cost way to explore dynamics over relatively long time periods. The advantage of reconstructed data may be particularly great when changes are infrequent, and potentially the only analytical option when contemporaneous data are unavailable.

The history of anti-smoking policies provides two motivating examples. First, the publication of the 1964 Surgeon General's Report on Smoking and Health is widely seen as an "information shock" that changed smokers' perceptions about the riskiness of smoking. However, there is relatively little contemporaneous data available to study its impact on smoking cessation.[2] Second, a more recent example is the impact on smoking cessation of the 1998 Master Settlement Agreement reached between the tobacco industry and state attorneys general. Reconstructed discrete-time hazard data offer opportunities to determine if the hazard of smoking cessation increased following these events. These opportunities seem to be expanding: in recent years, the number of cross-sectional and longitudinal surveys to include retrospective questions about the timing of smoking initiation and cessation has increased dramatically.[3]

Inaccurate recall of past behavior might mean that the research opportunities presented by reconstructed discrete-time hazard data are partly illusory. On one hand, contemporaneous self-reported smoking status is generally reliable (Patrick et al., 1994), and other studies find substantial agreement between contemporaneous and retrospective reports of smoking status (Machlin et al., 1989; Kenkel et al., 2003).[4]

---

[2] The Current Population Survey collected the first federal statistics on smoking prevalence in the United States in 1955. A second round of data collection occurred in 1966. Separately, the National Health Interview Survey first included questions on tobacco use in 1965.

[3] Retrospective questions about initiating and quitting smoking are included in the cross-sectional U.S. National Health Interview Survey and the Tobacco Use Supplements to the Current Population Survey. Retrospective questions about smoking are also included in recent waves of: three of the samples of the National Longitudinal Surveys Original Cohorts; the National Longitudinal Survey of Youth 1979; the Panel Study of Income Dynamics; and the Health and Retirement Study. Similar questions are included in the British Household Panel Survey, the German Socio-economic Panel Survey, the Russian Longitudinal Monitoring Survey, and the China Health and Nutrition Survey.

[4] The evidence on the reliability of contemporaneous reports of smoking status comes from studies that compare self-reports to biochemical markers of smoking (Patrick et al., 1994). Note that it is possible for retrospective reports of smoking status to be fairly accurate, as Machlin et al. (1989) and Kenkel et al. (2003) find, even when heaping in the reported date of cessation is common. Inaccurately recalling year of cessation by a small number of years over a long period of smoking results in only a small fraction of observations where smoking status is misclassified. The inaccuracy of a few years over a shorter period of smoking, however, will result in a more substantial fraction of misclassified smoking status.

On the other hand, retrospectively reported data frequently exhibit "heaping," a recall bias phenomenon resulting in abnormal concentrations of survey responses at certain values. This form of measurement error is observed in a variety of outcomes of interest such as duration of unemployment (Torelli and Trivellato, 1993; Abrevaya and Hausman, 1999), months attempting pregnancy (McLain et al., 2014), and year of smoking cessation (Forster and Jones, 2001).

[Figure 1. Distribution of Recalled Year of Cessation among Former Smokers, CPS TUS 2002]

Data from the 2002 Tobacco Use Supplement to the Current Population Survey (CPS TUS) provide a typical example of heaping in retrospective reports of smoking cessation. The survey included questions that identified both current and former smokers and their current and former smoking behaviors.[5] The Appendix provides additional information on the CPS TUS and the 2002 sample. Figure 1 shows the distribution of recalled year of cessation among former smokers with the characteristic "heaps" of responses at multiples of 5 years before the survey. Converting these responses and observed rates of smoking from retrospective reports of smoking initiation, Figure 2 shows the implied cessation rate over time. The data show pronounced heaping at 5 years intervals before the survey year of 2002. For example, the cessation rate in 1992, exactly 10 years before the survey, is almost 6% compared to less than 1.5% in 1991 and 1993. The cessation rate suggests similar heaping in 1987, 1982, 1977, and so on. In data from the 2003 TUS-CPS, the heaping is shifted forward one year, with much higher cessation rates in 1993, 1988, 1983, 1978, and so on.

[Figure 2. Observed Rate of Smoking Cessation, CPS TUS 2002]

The heaping problem is an example of a mismeasured dependent variable. This study is related to a line of applied econometrics research on similar measurement problems in labor and health economics. In the ordinary least squares context, classical measurement error in the dependent variable is fairly innocuous; it simply leads to less statistical precision in estimation (Hausman, 2001). In non-linear models, however, mismeasurement of the dependent variable leads to estimators that are biased and inconsistent. Meyer and Mittag (2014) review the bias in binary choice models. Poterba and Summers (1995) study misclassification errors in a multinomial logit model of employment status. Hausman, Abrevaya and Scott-Morton (1998) consider misclassification error in a probit model of job change. Kenkel, Lillard and

[5] Smokers attempting to reduce their consumption or quit completely commonly go weeks or months without smoking yet later resume smoking. In this analysis we focus on permanent cessation without resumption and, even during periods when smoking behavior may be less intense or paused altogether, we model the respondent as a smoker up until the reported permanent cessation.

Mathios (2004) apply the Hausman et al. (1998) approach to a probit model of smoking participation. Keane and Sauer (2009) incorporate misclassification error into a dynamic discrete choice model of female labor supply. In non-linear models with continuous dependent variables, Torelli and Trivellato (1993) and Abrevaya and Hausman (1999) consider measurement error in data on the duration of unemployment. Forster and Jones (2001) and Forster and Smith (2011) apply the Torelli and Trivellato (1993) approach to data on the duration of smoking and survival under slavery, respectively. A related line of research has investigated methods for smoothing heaped distributions for subsequent analysis (Heitjan and Rubin, 1990; Bar and Lillard, 2012; McClain et al., 2014).

This analysis contributes to the literature by proposing a model for the observed heaping behavior and evaluating the performance of several methods meant to account for recall error in a Monte Carlo simulation. We connect the heaping research in continuous-time hazard models (Torelli and Trivellato, 1993) to estimation of discrete-time hazard models (Allison, 1982; Jenkins, 1995) by linking recall error to misclassification in the cross-section (Hausman et al., 1998). In the context of smoking cessation, results suggest that the bias in estimated regression coefficients and marginal effects in a discrete-time hazard model are modest: less than 7% difference for estimated coefficients and 5% difference for marginal effects. Methods intending to account for such bias typically perform worse than the naïve model which ignores heaping, and several perform poorly across the board. Finally, we document the substantial bias that may still persist as a result of heaping in a simulation of a policy intervention. We believe these results provide guidance to researchers assessing analytical methods when faced with heaped data.

The following section reviews recall bias and discusses the impact of the resulting errors in cross-sectional and discrete-time hazard models. Section 3 presents a model of recall error and develops a method to simulate that model. Section 4 describes several methods suggested to account for recall error. Section 5 presents results for the estimated bias when recall error is ignored and compares the analytical methods' performance to account for this bias from the Monte Carlo analysis. Section 6 investigates a simulated policy intervention similar to the Surgeon General's Report to characterize the degree of bias one may encounter when using reconstructed hazard data for policy analysis. Section 7 makes concluding remarks.

2. A Maximum Likelihood Approach to Inaccurate Recall Due to Heaping

The functioning of human memory and its impact on the responses to survey questions has received a great deal of attention in cognitive psychology and epidemiology. A summary conclusion is that respondents' memories are often unreliable and their responses inaccurate. Memory biases related to retrospectively gathered data include reporting bias, the result of inaccurate reporting on whether an event occurred, and recall bias, the result of systematic error in the accuracy or completeness of recalled events (Last, 2001). Telescoping, the incorrect placement of an event within time, is a particular concern in hazard data as it directly impacts the dependent measure of interest; see, for examples, Bowers and Horvath (1984) and Beckett et al. (2001).

Misclassification Error in a Cross-Sectional Probit Model

In the reconstructed hazard data, recall inaccuracy results in misclassification in the dependent variable for the observed observations. In Figure 1, for example, the observations for 1991 include quits that are misclassified as non-quits, while the observations for 1992 include non-quits that are misclassified as quits. Here we review the maximum likelihood model of misclassification error in a binary choice model laid out in Hausman, Abrevaya, and Scott-Morton (1998) and expanded in Meyer and Mittag (2014).

Equation (1) describes an empirical demand function for smoking cessation. Let y_i^* be the latent variable giving the net benefits of smoking cessation as a function of observable determinant vector x_i and a random disturbance term ε_i:

$$y_i^* = x_i{}'\beta + \varepsilon_i \qquad (1)$$

The net benefits of smoking cessation y_i^* could be formally defined as the difference in lifetime utility from quitting and the lifetime utility from continuing to smoke (Becker and Murphy, 1988; Suranovic et al., 1999; Jones, 1999). In contrast to Equation (1), a standard approach in health economics is to estimate a model of smoking participation, and frequently in conjunction, a two-part model of the quantity of cigarettes smoked conditional upon smoking participation; see Chaloupka and Warner (2000) for a review. As Forster and Jones (2001) estimate and DeCicca, Kenkel and Mathios (2008) discuss in more detail, the addictive nature of smoking makes it important to model smoking initiation and cessation as distinct behaviors.[6]

[6] DeCicca et al. (2008) show that a myopic addiction model leads to distinct models of smoking initiation and cessation. The specification of the standard model of smoking cessation is correct only if smoking is not addictive: it implicitly imposes the testable restriction that the explanatory variables should have symmetric effects on initiation and cessation. Not surprisingly, in

The individual quits smoking when the net benefits of quitting are positive, so letting $\tilde{y}_i$ be the true response (here, true smoking cessation),

$$\tilde{y}_i = \begin{cases} 1 & if\ y_i^* > 0\,, \\ 0 & otherwise. \end{cases} \tag{2}$$

In the absence of misclassification, consistent estimation of β can be achieved via maximum likelihood. Instead, the observed dependent variable y_i takes the value of 1 if the individual *reports* smoking cessation in the period and 0 otherwise. Let α_0 be the probability that a true 0 (smoking continuation) is misclassified as a 1 (cessation), and let α_1 be the probability that a true 1 is misclassified as a 0 for all responses:

$$\begin{aligned} \alpha_0 &= \Pr(y_i = 1 |\, \tilde{y}_i = 0) \\ \alpha_1 &= \Pr(y_i = 0 |\, \tilde{y}_i = 1) \end{aligned} \tag{3}$$

Meyer and Mittag (2014) refer to this set of assumptions as "conditionally random" because misclassification is independent of the covariates and conditional on $\tilde{y}_i$. The expected values of the true smoking cessation,$\tilde{y}_i$, and observed smoking cessation, y_i, are

$$E(\tilde{y}_i|x_i) = \Pr(\tilde{y}_i = 1|x_i) = F(x_i{}'\beta) \tag{4}$$

$$E(y_i|x_i) = \Pr(y_i = 1|x_i) = \alpha_0 + (1 - \alpha_0 - \alpha_1)F(x_i{}'\beta) \tag{5}$$

for a distributional function F. Typically, we are interested in the partial derivative of Equation (4) with respect to a particular independent variable x_{ij}; for example, we might be interested in the effect of a marginal change in the price of cigarettes on the probability of true smoking cessation. However, comparing the partial derivative of Equation (4) with the partial derivative of Equation (5) shows that with misclassification error, the estimated marginal effects of interest will be biased towards zero:

$$\frac{\partial \Pr(\tilde{y}_i = 1|x_i)}{\partial x_{ij}} = f(x_i{}'\beta)\beta_j \tag{6}$$

$$\frac{\partial \Pr(y_i = 1|x_i)}{\partial x_{ij}} = (1 - \alpha_0 - \alpha_1)f(x_i{}'\beta)\beta_j \tag{7}$$

Under the "conditionally random" assumptions and with known F, Hausman et al. (1998) give the full log-likelihood as

their empirical analysis DeCicca et al. reject the hypothesis that smoking is not addictive. DeCicca et al. provide some discussion of how the approach could be extended to allow for rational addiction or different degrees of addiction. These extensions are beyond the scope of this paper, but in the conclusion we discuss how future work might usefully explore duration dependence, which is related to the degree of addiction.

$$\mathcal{L}(\alpha_0, \alpha_0, \beta) = \frac{1}{N}\sum_{i=1}^{N} \{ y_i \ln\big(\alpha_0 + (1-\alpha_0-\alpha_1)F(x_i{}'\beta)\big) + (1-y_i)\ln(1-\alpha_0-(1-\alpha_0-\alpha_1)F(x_i{}'\beta))\}. \quad (8)$$

with N as the number of responses. Under this approach, the misclassification probabilities are estimable parameters: significance tests on α_0 and α_1 provide tests for misclassification error.

Hausman, Abrevaya and Scott-Morton (1998) point out that identification of the model parameters stems from the nonlinearity of F, so that in the linear probability model the parameters are not separately identified. They discuss an additional monotonicity condition required for identification because, for a symmetric F, like the normal or logistic, the MLE estimator cannot distinguish between the parameter values $(\alpha_0, \alpha_1, \beta)$ and $(1-\alpha_0, 1-\alpha_1, -\beta)$. Imposing the monotonicity assumption that the sum of the misclassification probabilities must be less than one $(\alpha_0 + \alpha_1 < 1)$ rules out the latter set of parameter values.

To shed light on the empirical importance of even modest misclassification, Hausman et al. (1998) report Monte Carlo simulation results. They consider symmetric misclassification probabilities $(\alpha_0 = \alpha_1)$ of 2%, 5%, and 20%. Even with $\alpha_0 = \alpha_1 = 0.02$, naïve probit estimates are underestimated by 15% to 25%. When misclassification error is greater, the naïve probit estimates are underestimated by 62% to 81%. In contrast, their modified "misclassification MLE model" results, based on Equation (8), yield estimates of α_0 and α_1 (restricted to be equal) and β that are very close to the simulation parameters.

Recall Error in the Discrete-Time Hazard Setting

The relationship between recall error in the discrete-time hazard setting and misclassification in the cross-sectional setting begins with the reconstruction process to obtain the hazard data. If the recalled year of cessation is later (in a calendar year sense) than the actual year of cessation, the data observes the respondent at risk for too many periods and additional observations for the respondent are constructed erroneously; if the recalled year of cessation is earlier than the actual year, the process considers the respondent at risk for too few periods and the respondent is observed in the reconstructed data less than the true number of years. In the former case, misclassification occurs in the year of true cessation – a true cessation (1) is coded as a continuation (0); in the latter, misclassification occurs in the heaped year – a true continuation is misclassified as a cessation. As a result of the reconstruction process, it is likely that only the last few at risk periods of an individual's discrete-time hazard would be misclassified. However, depending on average duration and the degree of recall inaccuracy, even a fraction of the bias found by Hausman, Abrevaya, Scott-Morton (1998) would be a serious concern.

3. A Model of Recall Error and Monte Carlo Methods

We propose a simple model of recall error which generates heaping consistent with observed data. Next, we conduct a Monte Carlo simulation analysis to assess the impact of inaccuracies due to heaping on estimated coefficients and marginal effects. We then compare several approaches intended to account for the heaping and whether any observed bias is diminished.

"Random Heaping"

Our proposed model of recall error, herein referred to as "random heaping," is similar to Meyer and Mittag's (2014) model of "conditionally random" misclassification in that recall error is independent of the covariates and equal across all respondents. The "random heaping" model posits that individuals correctly recall if cessation occurred in any year which is a multiple of 5 years ago, but that one-third of all people who quit in years which are not a multiple of 5 years ago randomly heap to the nearest 5-years-ago period. For example, there is a probability of one-third that respondents who quit 8, 9, 11, or 12 years ago will report quitting 10 years ago.[7]

"Random heaping" is based on a simple assumption of rounding to the nearest observed heaped value. The model is consistent with digit preference (Ridout and Morgan, 1991) and reduced cognitive effort via overall mindlessness (Langer, 1992) or satisficing, expending only the amount of effort necessary to make an acceptable or satisfactory survey response (Krosnick, 1999). With an identifiable set of heaped values, "random heaping" is applicable to the aforementioned labor market and health behavior outcomes. Further, the general model of "rounding to the nearest heaped value" may be expanded; we consider two such expansions in Appendix 2.

[7] We apply one exception to the "random heaping" rule: respondents do not exhibit recall error if the true duration of smoking is less than 3 years. In such a case, recall error may be great enough such that cessation would be reported prior to initiation.

Monte Carlo Methods

Our Monte Carlo simulation design mimics the reported data on smoking cessation collected retrospectively from a survey that asks current and former smokers “About how many years ago did you quit smoking completely?" Following the setup of Allison (1982) and Jenkins (1995), we begin with the traditional latent variable approach and model that an individual “quits” when the latent index first exceeds zero. The individual remains in the at-risk sample for every period until the quit occurs or is right censored.

To present a realistic yet parsimonious model of smoking cessation, we assume that the latent index that determines smoking cessation is a function of five covariates providing both time-varying and time-invariant variation: the real price of cigarettes, age, gender, education and household income. These covariates are considered by Kenkel, Lillard and Mathios (2004) in a study of smoking participation using retrospectively collected data. To capture the correlations between these variables, we generate the initial observations of each respondent by bootstrap sampling with replacement from the 2002 CPS TUS data and simulating forward. The price of cigarettes evolves over time by adding a random, uniformly distributed variable over [-0.25, 0.50] in each period. Age increases by one each year, and education is calculated as the minimum of the observed total years of education and age minus six. Female and household income in 2002 (in $10,000s) are time-invariant. See Appendix 1 for a description of the 2002 CPS TUS and the covariates above. Right censoring occurs after 1999.

The simulation values for the index coefficients were taken from a naïve probit estimation of the 2002 CPS TUS data and inflated up to offset the expected attenuation. Including a quadratic time trend, the latent dependent variables is

$$\begin{aligned} y_{it}^{*} &= 0.2\, price_{it} + 0.02\, age_{it} + 0.03\, female_{i} + 0.08\, education_{it} + 0.008\, income_{i} \\ &\quad + 0.02\, year_{it} - 0.003\, year_{it}^{2} - 3.8 + \varepsilon_{it} \end{aligned} \tag{9}$$

where $year_{it}$ is measured from the earliest year of observed initiation in the sample (here, 1954) and ε_{it} is distributed as a standard normal variable.

From the actual year of cessation, “random heaping” is induced according to the model described above: one-third of all respondents inaccurately recall to the nearest heaped year, those years that are multiples of five years before the survey year. Only respondents smoking more than two years were heaped. For the 2002 survey year, the heaped years thus correspond to 1957, 1962, 1967, and so on, up to 1997.

In addition to evaluating the impact of heaping on the estimated coefficients, we also evaluate its impact on the estimated marginal effects. The Hausman et al. (1998) model gives that marginal effects will be biased by a constant factor in the cross-sectional results and finds even small misclassification results in substantial bias. In order to present a consistent measure of bias in the marginal effects, we compare the average marginal effects for first observation of each respondent in the 2002 CPS TUS, from which we bootstrap sample, to the average marginal effects for the first observation in each of the simulated samples.

4. Methods to Correct for Inaccurate Recall Due to Heaping

Reduced Form and Subsampling Methods

The first method to account for heaping is the addition of an indicator variable to the discrete-time hazard model indicating whether the observation falls in a heaped year. Intuitively, by controlling for the increase in the unconditional hazard for those periods, the indicator variable will reduce any bias resulting from heaping. Torelli and Trivellato (1993) and Forster and Jones (2001) consider this approach in the context of a continuous time duration model, inspired initially by Hujer and Schneider's (1989) analysis of labor market mobility which included indicator variables. Kraus and Steiner (1996) evaluate the inclusion of heaping indicators in a discrete-time hazard model and found that the indicators picked up selection effects, finding that indicators are appropriately used under a generalized model of heaping. Kraus and Steiner's (1996) findings are used as justification in Hujer et al. (1997) and for others' use of heaping indicators when faced with heaped data.

The second method to account for heaping subsamples the respondents, eliminating all respondents whose recall of year of cessation is a heaped year. Herein we refer to this approach as analysis on the "decimated sample." This approach aims at eliminating all respondents which may be contributing measurement error to the regression model. Necessarily, the approach simultaneously distorts the rate of observed smoking cessation.[8] To deal with a similarly observed form of measurement error in reported birth weights, Barreca, Lindo, and Waddell (2011) expand on a novel approach by Barreca, Guldi, Lindo, and Waddell (2011), proposing an approach similar to decimation. Their focus is on the implications for regression discontinuity designs where birth weight is the running variable and is disproportionately represented at multiples of round numbers, especially for children of lower socioeconomic status mothers. They suggest that a straightforward approach to deal with the problem is what they call a "donut RD," "dropping observations at data heaps" (Barreca, Lindo, and Weddell, 2011).

Misclassification MLE Model

The use of a likelihood correction to account for inaccurate recall naturally follows from the misclassification models discussed above. Under this approach, Hausman, Abrevaya, and Scott-Morton's (1998) misclassification probit is applied period-by-period, allowing the misclassification probabilities to detect the rate and impact of heaping. Under the "random heaping" rule described above, the pattern of observations reveals exclusion restrictions for misclassification probabilities. In the two years prior to a heaped year, only true cessation can be misreported because inaccurate recall falsely generates observations of continued smoking during those years. As such, $\alpha_1 > 0$ in those years and $\alpha_0 = 0$. Conversely, in the heaped year, all true cessation is observed correctly, yet some cessation that would occur in the two years after the heaped year are falsely observed, such that $\alpha_0 > 0$ and $\alpha_1 = 0$. In the two years after a heaped year, no false smoking continuation or cessation is observed since any inaccurate recall has heaped false cessation to the heaped year. These exclusion restrictions lead to the following misclassification probabilities in a discrete-time hazard model of "random heaping":

$$(\alpha_0, \alpha_1) = \begin{cases} (\quad 0 \quad , \alpha_{1N} > 0\,) & 2\ \textit{years prior to heaped year,} \\ (\quad 0 \quad , \alpha_{1N} > 0\,) & 1\ \textit{year prior to heaped year,} \\ (\,\alpha_{0H} > 0\,, \quad 0 \quad) & \textit{Heaped year,} \\ (\quad 0 \quad , \quad 0 \quad) & 1\ \textit{year after heaped year,} \\ (\quad 0 \quad , \quad 0 \quad) & 2\ \textit{years after heaped year.} \end{cases} \tag{10}$$

Because inaccurate recall for true cessation two years prior the heaped year generates a misclassified continuation in both two years prior and one year prior to the heaped year, we allow α_{1N} to be a function of whether the observation is one or two years prior to a heaped year. Further, to account for the growth in

[8] From the perspective of year of cessation recall, "decimation" can be viewed as producing an endogenously stratified sample from the population. The strata corresponding to heaped quit years are missing from that sample though all non-heaped years are perfectly observed. The statistical literature has developed estimation techniques to analyze stratified data and Cosslett (1993) provides a thorough review. The general solutions are (i) supplementing the sample with exogenous information, and (ii) jointly estimating the marginal probabilities of the outcomes as part of a pseudo-likelihood. Meyer and Mittag (2014) also suggest these solutions in their analysis of misclassification in binary choice models.

the rate of cessation and heaping over time, we allow α_{0H} to be a linear function of the year of observation.

An additional concern in the estimation of the Hausman, Abrevaya, Scott-Morton (1998) misclassification MLE model is the monotonicity condition that the sum of the misclassification probabilities impacting any observation must be less than one. Based on our exclusion restrictions on the αs under "random heaping," we need simply ensure that each α is less than one. Therefore, estimation is carried out using the logistic function which naturally satisfies the monotonicity assumptions:

$$\alpha = L(z) = \frac{e^z}{1+e^z} \tag{11}$$

Put together, the log likelihood function for the misclassification MLE model maximized is

$$\begin{aligned}\mathcal{L}(\lambda,\mu,\beta) = \frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\{\, & y_{it}\ln\big(\alpha_0 + (1-\alpha_0-\alpha_1)F(x_{it}{}'\beta)\big) \\ & + (1-y_{it})\ln(1-\alpha_0-(1-\alpha_0-\alpha_1)F(x_{it}{}'\beta))\}\end{aligned} \tag{12}$$

with

$$\alpha_{0H} = L(\lambda_0 + \lambda_1\, year_{it})$$
$$\alpha_{1N} = L(\mu_{2\ years\ before\ a\ heap} + \mu_{1\ year\ before\ a\ heap})$$

where N is the number of respondents, T_i is the total number of observed periods for respondent i, α_0 and α_1 operate according to Equation (10) above, and λ and μ are the coefficient vectors in the logistic functions for α_{0H} and α_{1N}, respectively.

Multiple Imputation

A fourth method to account for heaping utilizes multiple imputation to overcome inaccuracies in recall of the year of cessation. Heitjan and Rubin (1990) implement a multiple imputation approach to heaped age data.[9] Along a similar vein, Torelli and Trivellato (1993) and Forster and Jones (2001) implement a smoothing approach which randomly redistributes the observations at heaped values within the interval

[9] Heitjan and Rubin (1999) refer to their data as "coarse" to suggest the responses were not collected at a "fine" level of detail. In contrast to the multiple imputation method, which attempts to identify a "finer" response than the heaped year, we additionally estimated an approach which "coarsened" all of the data, such that the unit of analysis is the period of five years around a heap. The goal of this approach was to eliminate the effect of heaping by eliminating heaps and instead focusing on cessation over wider intervals, made simple by the discreteness of the hazard function. However, the coarsening of the data eliminates a significant number of observations. The results (available upon request) for this method showed it produced significantly biased results: the coefficients were biased on average between 12% and 113% and the marginal effects between 94% and 369%. In the policy analysis experiment, the bias was several times larger than even that top number.

around the heap. Multiple imputation improves upon this method by reassigning heaped values within interval as predicted by the observed characteristics among respondents with non-heaped values. Operationally, all respondents who report a heaped year have their recall set to missing, and multiple imputation via truncated regression imputes the estimated recall within the five year period around the heaped year. Stata 13's truncated regression imputation is a parametric imputation method based on the asymptotic approximation of the posterior predictive distribution of the missing data, and here utilizes the joint variation in recall, age, gender, education, income, and reported duration (in years) of smoking. Predicted recall from the imputation is rounded to the nearest whole year and the resulting discrete-time hazard data is reconstructed from the imputed recall. This analysis uses $m = 5$ imputations per respondent and averages the estimates (Rubin, 1987) from each imputation to form the single multiple imputation estimate per simulated sample.

Naïve Models

Of course, to quantify the degree of bias created by inaccurate recall, we estimate a naïve probit model that ignores the heaping effect from inaccurate recall. Indeed, in duration analyses with heaping, the naïve model has generally found a small impact from heaping. Torelli and Trivellato (1993) note that "[w]hile it is clear that parameter estimates of the duration model given by [the specified model of heaping] and [the naïve model ignoring the heaping effect] do not differ appreciably, in our opinion this fact should not be overemphasized." Nevertheless, Forster and Jones (2001) and Forster and Smith (2011) again find only modest differences between the naïve and modeled approaches to account for heaping.

Finally, we provide results from the estimation of a linear probability model on the discrete-time hazard data. The appeal of linear probability models and the ease of their interpretation has made them widespread, including their application to hazard data. Margolis et al. (2014) apply a discrete-time linear probability hazard model to assess whether smoking cessation accounts for the disparity in outcomes between patients receiving different treatment for coronary artery disease. In the cross-sectional binary choice observation, the linear probability model still suffers in the face of misclassification error as shown by Meyer and Mittag (2014).

5. Simulation Results

Table 1 provides the Equation (9) coefficient estimates from the different methods to account heaping in the simulated data. To focus on the degree of bias we report differences between the mean estimates rather than the standard deviation of the results of the models.

[Table 1. Coefficient Estimates from Simulated Smoking Hazard Data with Respondent Recall of "*How Many Years Ago Did You Quit Smoking?*" Randomly Heaped to *5*s and *10*s]

In Table 1, the coefficient results for the probit estimated on the non-heaped ("true") data are biased under one-half of one percent. This confirms that the simulation was sufficiently powered to correctly return the simulation coefficients. The naïve probit model on average yielded estimated coefficients which were only modestly different from the true values: the estimated coefficients were less than 5% smaller than the true values and the time trend coefficients were less than 10% smaller than the true values. None of the other methods represent a clear improvement over the naïve probit and some perform much worse. Compared to the naïve probit model results, including a heaping indicator did not uniformly increase or decrease the average bias. The multiple imputation model was on average more successful at identifying the time trend but less successful at estimating the covariates of interest without bias. The probit on the decimated sample and the misclassification MLE yield estimated coefficients that are more biases, sometimes substantially so, than the estimates of the naïve probit model.

[Table 2. Additional Coefficient and Parameter Estimates from Simulated Smoking Hazard Data with Respondent Recall of "*How Many Years Ago Did You Quit Smoking?*" Randomly Heaped to *5*s and *10*s]

Table 2 gives the additional coefficient and parameter estimates from the probit model including a heaping indicator and the misclassification MLE model. As expected, the coefficient on the indicator for heaped years was positive because it models an unconditional increase in the hazard of cessation in heaped years. The estimated constant and year coefficient in the function for α_{0H} suggested that α_{0H} grows over time, from 2% to 15% between 1957 and 1997. The most direct measure of the misclassification MLE model's ability to recover the simulation parameters is α_{1N} two years prior to a heaped year. The model over-estimated this probability of heaping, 33.33% in the simulated data, at greater than 40%. The estimated difference in misclassification probabilities between the two years leading up to a heaped year was small.

[Table 3. Marginal Effect Estimates from the First Year of Observation of Each Respondent in the Simulated Smoking Hazard Data with Respondent Recall of "*How Many Years Ago Did You Quit Smoking?*" Randomly Heaped to *5*s and *10*s]

Although coefficient estimates in a probit model are insightful to the direction and statistical significance of a covariate's impact, we also present comparisons of the methods' average marginal effects. Marginal effects would be of greater interest to policy makers evaluating methods to increase permanent cessation. Table 3, focusing on estimated marginal effects in the first year of observation for all respondents, again shows that the naïve probit model on average returned the least biased results: less than 4% for all covariates. This bias is significantly smaller than the bias predicted by Equations (6) and (7) above. The marginal effects estimates from the linear probability model were predictably less biased than the coefficients, but the bias remained in excess of 25%. The probit model with a heaping indicator again exhibited greater bias than the naïve probit in general but overall performed better than the remaining methods. The probit model estimated on the decimated sample varied in the average observed bias, ranging from 0.22% below the true value to more than 35% greater than the true value. The results from the multiple imputation approach also varied but to a smaller degree: from 0.21% below the true value to more than 10% above. The misclassification MLE exhibited a moderate amount of bias in the estimated marginal effects, 10% to 26% above the true value.

Across the coefficient and marginal effect estimates, the naïve probit model most often results in the least amount of bias, and that degree of bias is modest: less than 7% difference for estimated coefficients and 5% difference for marginal effects. The models with the greatest theoretical appeal, the misclassification MLE and multiple imputation approaches, do not yield consistent results.

6. Evaluation of a Policy Intervention

A primary goal of using retrospective behavioral data from surveys is the evaluation of past policies, or information changes. For smoking, an important informational change occurred in 1964 with the release of the first report of the Surgeon General's Advisory Committee on Smoking and Health. The report laid out the negative health consequences of smoking and it was one of the top news stories of 1964 (CDC, 2014). To simulate the impact of heaping on such a policy evaluation, we expand the model of smoking cessation above to include a simulated policy intervention. The intervention is modeled by including an indicator variable that evaluates differently before and after a date. One can easily imagine circumstances under which the year of implementation of the policy of interest may align or misalign with an observed heap in the retrospectively collected data, so we also simulate the policy intervention starting in each year around a heaped year.

To both estimate the extent of bias we might expect under "random heaping" in a policy analysis paradigm, and understand how any bias might vary according the year of the policy implementation relative to a heaped year, we simulated a model of smoking cessation with a hypothetical policy intervention starting each year between 1980 and 1984, the five years around the 1982 heap. In the latent index, the policy year switch takes a value of 1 beginning in the year of implementation and 0 in all years prior. For simulation, the coefficient on the policy year switch is 0.1, such that the policy intervention increases the latent index of cessation by an amount equal to a $0.50 increase in the real price of cigarettes.[10]

[Figure 3. Bias in the Estimated Marginal Effect of a Simulated Policy Intervention, 1980–1984]

Figure 3 presents the bias in the estimated marginal effect of the simulated policy intervention for each method when the intervention is implemented each year between 1980 and 1984. Across all models, the average bias was considerable, often much larger than the moderate levels found in the "random heaping" analysis. The bias that policy analysis would be subjected to, as estimated by the average bias of the results of the naïve probit model, ranged from approximately 22% to 160% in absolute magnitude. The largest positive bias occurred the year of the heap, which has many false quits, and the largest negative bias in the year just after the heap. The misclassification MLE model exhibits bias in the opposite direction as the naïve probit model but on a similarly large scale. The probit model with the heaping indicator is closest to the true marginal effect, though on average still considerably biased: more than 20% in the years furthest from the heap.

---

[10] A $0.50 increase in the real price of cigarettes is an increase of approximately one-third for the period 1980 to 1984.

7. Discussion

The ability to use information about past behavior collected retrospectively in surveys to estimate behavioral determinants or policy impacts may be hindered by inaccuracies in survey responses. Recall error, and the resultant "heaping" in particular, is a well-documented survey-response phenomenon exhibited across a variety of applications. As a form of measurement error, the inaccuracy has been demonstrated both theoretically and empirically to bias estimates in a cross-sectional setting as misclassification error, as well as in some duration models. This paper proposed a model of observed recall error, "random heaping," the rounding of the recalled year of smoking cessation to the nearest multiple of 5 years before the survey year, and evaluated several methods which intended to account for the impact of heaping.

The results of a Monte Carlo simulation of "random heaping" confirmed that recall error leading to heaping in "years ago quit," and the derived misclassification error, produces biased results if the heaping is ignored. Although the average bias in the estimated coefficients and marginal effects in a model of the determinants of smoking cessation were less than 7% and 5%, respectively, the bias reached tremendous proportions in the context of a simulated policy intervention: in one scenario, the marginal effect of the policy intervention was biased, on average, by more than 160% in absolute magnitude. Although the results of the policy intervention are specific to the simulation here, taken with the earlier results, recall error was indeed found to induce bias in both estimated coefficients and marginal effects. Depending on the setting and model, however, that bias may be quite small or quite large. Future work may investigate which and how the characteristics of covariates are deterministic of the degree of bias in order to recommend or warn against the use of reconstructed hazard data for analysis.

With the existence of the bias confirmed, we evaluated several methods suggested by the literature which intended to account for the inaccuracy in recall to provide consistent results. Strikingly, in the model of the determinants of smoking cessation, the "cure" was often worse than the "disease," and the naïve probit model typically performed the best; the analyses ignoring the recall error on average produced the lowest degree of bias for each of the covariates' estimated coefficients and marginal effects. Perhaps we should not be surprised by this result. Heitjan and Rubin (1990), Torelli and Trivellato (1993), and Forster and Jones (2001) all found that results from the naïve model in their analyses did not differ substantially from those of the "theoretically more satisfactory" models (Heitjan and Rubin, 1990). Torelli and Trivellato (1993), in fact, find the naïve model performed better than the *ad hoc* approaches as we do. When the results of the simulated policy analysis are considered, the leased biased approach seems to be the inclusion of a heaping indicator. Although the average marginal effects for that model were biased by 5% to 20% in magnitude in the simulation, the naïve probit model returned average marginal effects that were biased 22% to 162%. Between the two approaches, neither method is consistently less biased than the other, and future work should consider using an even larger window of start dates for the policy intervention to determine if these results are robust beyond our experiment.

Despite their theoretical appeal, neither the misclassification MLE nor multiple imputation approach reduced the bias observed in the naïve probit model. For both methods, the exacerbated bias may a result of whether a particular person-year is observed more so than what value that observation takes. Recall that observations in the two calendar years after a heaped year contain only correctly classified observations, yet the quit rate does not take its true value as a result of changes to the numerator and denominator of the rate in those years. Although the misclassification MLE model is the correct model for the behavior that is observed by the econometrician, the model does not account for the dependence of the sample number of observations on heaping. When applying multiple imputation, the imputed value of "years ago quit" used to reconstruct the hazard data only results in the "true" data when the imputation model perfectly predicts the true year of cessation. Otherwise, imputation generates "falsely observed" and "falsely unobserved" observations compared to the true data. When the false observations generate misclassification error, we know from Hausman et al. (1998) the attenuation bias in the marginal effects will not be averaged out through multiple imputation. How the false observations may counter this effect, however, should be explored in future work, potentially by incorporating a selection model of observation related to the heaping process.

Although we attempted to simulate a model which closely approximates the true behavior underlying the CPS TUS data, simulations necessarily deviate from the complexity of the real world. Neither our "random heaping" model nor the models discussed in Appendix 2 fully and accurately capture the behavioral process underlying responses to "how many years ago did you quit smoking?" As discussed above, in the absence of the externally validating data suggested by Meyer and Mittag (2014), or other surveys questions which may suggest a respondent's pattern of heaping (Manski and Molinari, 2010), future work should consider the joint model of recall inaccuracy and hazard. Such a case may then allow the modified MLE approach to provide consistent estimates. To attempt to reduce the bias from a survey design approach, research into whether survey questions starting with "How long…" rather than "About how long…" might induce respondents to provide more accurate responses. Research on "demand effects" has shown that survey structure does influence survey responses.

Another direction for future work is to extend the discrete time hazard model to allow for duration dependence. The addictive nature of smoking suggests that the probability of quitting might fall with the duration of the habit, if longer durations are associated with a larger stock of addictive capital (Becker and Murphy, 1988). The discrete-time hazard model can flexibly allow for the probability of quitting to vary with duration, and does not impose positive or negative duration dependence. While we think that this makes the discrete-time hazard model an attractive approach to study smoking cessation, it will again be important to consider the bias induced by recall error.

Appendix 1. The Tobacco Use Supplement to the Current Population Survey

In February 2002, the Current Population Survey collected the Tobacco Use Supplement (CPS TUS), an in-depth survey of tobacco use sponsored by the National Cancer Institute. The CPS TUS is the primary source of U.S. national and state-level statistics on smoking, use of tobacco products, attitudes about tobacco use, and tobacco policies. Current and former smokers are identified according to their response to “Have you smoked at least 100 cigarettes in your entire life?” (“2001-2002 CPS Tobacco Use Supplement Items Booklet (with Skip Pattern),” 2004). Smoking initiation is identified via the question “How old were you when you started smoking cigarettes fairly regularly?” If the respondent reported no longer smoking, recall of smoking cessation is collected via the question “About how long has it been since you completely stopped smoking cigarettes?”

From the 2002 wave of the CPS TUS, we use the retrospectively collected data for 12,093 self-identified current and former smokers who initiated smoking during or after 1954. Respondents remain in the at-risk sample until they quit or are right-censored after 1999, resulting in 259,631 person-years of observation. A total of 5,620 or 46.47% of respondents indicate completely stopping cigarette smoking during the study period. The average annual quit rate is approximately 2%.

The explanatory variables in a discrete-time hazard model of smoking cessation using the CPS TUS sample are the real price of cigarettes, age, gender, education, and household income as well as a quadratic time trend. As in the simulated data, education is recoded to be the standard years of education up to the observed total years of education using the minimum of the observed total years of education and the respondents’ age minus six. This recoding of education is important since a great deal of continuation and cessation takes place during teenage and young adult years and analysis shows improved model fit using the recoded variable. The historical real price of cigarettes is merged to the CPS TUS data based on the year and the respondent’s state of residence in 2002, the latter of which introduces some measurement error in assigned prices due to cross-state movers. Household income is measured in $10,000s as of the 2002 and so should be viewed as a proxy for the respondent’s household income in all previous years.

[Table A1. Summary Statistics for the Sample from the Tobacco Use Supplement to the 2002 Current Population Survey]

Table A1 provides summary statistics for the CPT TUS sample. The current and former smokers in the sample began smoking in their mid-teens to early twenties. Those respondents who are observed to quit smoking in 1999 or earlier do so after an average of nearly 17 years of smoking, at an average age of 34 years. In the hazard data, respondents face an average real price of cigarettes of $1.94, and that price ranges from $1.10 to $3.80.

[Table A2. Coefficient and Average Marginal Effect Estimates from Discrete-Time Hazard Analysis of Smoking Cessation Reconstructed from the Tobacco Use Supplement to the 2002 Current Population Survey]

Table A2 presents results of a discrete-time hazard model estimated using the 12,093 person-year observations in the 2002 CPS TUS sample of 12,093 current and former smokers. The model is statistically significant (the *p*-value of the likelihood ratio χ^2 statistic is less than 0.0001) as are all of the covariates of interest and time trend (*p*-value less than 0.01 for all covariates except *Female*, for which the *p*-value is less than 0.10). The estimated coefficients are scaled up by 40% to 100% in order to be consistent with attenuation bias of 30% to 50% in a naïve probit model and within the range suggested by Hausman et al. (1998) for the simulated misclassification error.

Appendix 2. Robustness to Alternative Forms of Heaping

Because the true process of recall error is intrinsically more complex than "random heaping," we evaluate whether the conclusions drawn from the Monte Carlo simulations above hold under alternative models of heaping. The following two alternatives by no means explain the complete process of recall error but investigate specifically if the earlier results are robust to misspecification of the levels of the misclassification probabilities, and misidentification of "heaps" in the recall data.

"*Uneven Heaping*": *More to 10s than 5s*

A review of Figure 1 shows that respondents' propensity to inaccurately report year of cessation is likely different for those years which are multiples of 10 years prior to the survey (e.g., 10, 20, 30 years ago) compared the midpoints of those heaps (e.g., 5, 15, 25, 35 years ago). Under a model of "uneven heaping", the misclassification around multiples of 10 years ago is greater than that around the midpoints of those heaps: $\alpha_{1N-10s} > \alpha_{1N-5s}$. "Random heaping," in contrast, assumes an equal probability of recall error around 5 year intervals: $\alpha_{1N-10s} = \alpha_{1N-5s}$. To understand the robustness of the "random heaping" results to misspecification of the level of misclassification probabilities we simulate data for which respondents are twice as likely to heap to multiples of 10 years ago than the midpoints ending in "5": $\alpha_{1N-10s} = 2\,\alpha_{1N-5s}$; the implementation of the Monte Carlo simulation and implementation of the methods otherwise remains the same as for "random heaping."

[Table A3. Marginal Effect Estimates from the First Year of Observation of Each Respondent in the Simulated Smoking Hazard Data with Respondent Recall of "*How Many Years Ago Did You Quit Smoking?*" Randomly Heaped More to *10*s than *5*s]

Table A3 presents the summary statistics of the average marginal effects of the covariates from each of the analytical methods in the first year of observation under the misspecified model of "uneven heaping." As under the "random heaping" model, the naïve probit model on average exhibited the least bias. The magnitude of that bias is only slightly larger than under the "random heaping" model, and still less than 5%. The probit model with heaping indicator was again, on average, more biased than the naïve probit model. The remaining models exhibited bias on par with that seen under "random heaping."

"*Calendar Decade Heaping*"

A review of Figure 1 also shows a potential heap in the years of calendar decades prior to the survey year (e.g., 1970, 1980, and 1990). Under a model of "calendar decade heaping" we simulate data for which respondents who heap in the year before, year of, and year after a calendar decade (e.g., 11, 12, and 13 years ago corresponding to 1989, 1990, and 1991) are equally likely to heap to the calendar decade as the multiple of 5 year heap under the "random heaping" rule.[11] As above, the implementation of the Monte Carlo simulation and implementation of the methods otherwise remains the same as for "random heaping." A review of the results under "calendar decade heaping" will provide evidence on the robustness of the "random heaping" results to misidentifying or not exhaustively identifying heaped values in respondent recall.

[11] For example, one-third of respondents who quit smoking 8 or 9 years ago are expected to inaccurately report quitting as 10 years ago; one-sixth of respondents who quit 11 years ago are expected to report quitting as 10 years ago and one-sixth are expected to report quitting as 12 years ago; one-sixth of respondents who report quitting 12 years ago are expected to report quitting as 10 years ago; one-sixth of respondents who quit 13 years ago are expected to report quitting as 12 years ago and one-sixth are expected to report quitting as 15 years ago; finally, a full one-third of respondents who quit 14, 16, and 17 years ago are expected to report quitting as 15 years ago.

[Table A4. Marginal Effect Estimates from the First Year of Observation of Each Respondent in the Simulated Smoking Hazard Data with Respondent Recall of "*How Many Years Ago Did You Quit Smoking?*" Randomly Heaped to *5*s, *10*s, and *12*, *22*,… Representing Calendar Decades]

Table A4 presents the summary statistics of the average marginal effects of the covariates from each of the analytical methods in the first year of observation under the misspecified model of "calendar decade heaping." The average bias generated by the analytical methods under "calendar decade heaping" is often less than under "random heaping." This result is likely due to the reduced degree of recall error: respondents who heap to a multiple of 5 years ago or a calendar decade year must be, on average, closer to a heaped year compared to the "random heaping" model, which generates fewer misclassified observations. Those quitting in a calendar decade year have a reduced amount of any recall error. Nevertheless, the general extent and pattern of bias remains consistent across models under "calendar decade heaping" as under the other two forms of heaping: the naïve probit model most consistently produces the smallest bias, under 3%.

Discussion

Together these analyses suggest that our finding of the superior performance of the naïve probit model under "random heaping" is robust to certain misspecifications.

References

2004. 2001-2002 CPS Tobacco Use Supplement Items Booklet (with Skip Pattern). National Cancer Institute, Bethesda, MD, and Centers for Disease Control and Prevention, Atlanta, GA, last viewed October 27, 2014, at http://appliedresearch.cancer.gov/archive/tus-cps/tuscps_english_2001-02.pdf

2011. Selected Milestones in Tobacco Research, Promotion and Surveillance. Centers for Disease Control and Prevention, Atlanta, GA, last viewed October 28, 2014 at http://www.cdc.gov/nchs/nhis/tobacco/tobacco_history.htm

2014. History of the Surgeon General's Reports on Smoking and Health. Centers for Disease Control and Prevention, Atlanta, GA, last viewed on October 27, 2014, at http://www.cdc.gov/tobacco/Data_statistics/sgr/history/index.htm.

2014. Stata 13 and Stata Multiple Imputation Reference Manual Release 13. StataCorp LC, College Station, TX.

Allison PD. 1982. Discrete-time methods for the analysis of event histories. *Sociological Methodology* 13: 61-98.

Abrevaya J, Hausman JA. 1999. Semiparametric estimation with mismeasured dependent variables: an application to duration models for unemployment spells. *Annales d'conomie et de Statistique/Annals of Economics and Statistics* 55/56: 243-275.

Bar HY, Lillard DR. 2012. Accounting for heaping in retrospectively reported event data – a mixture-model approach. *Statistics in Medicine* 31: 3347-3365.

Barreca A, Guldi M, Lindo J, Waddell G. 2011. Saving babies? revisiting the effect of very low birth weight classification. *Quarterly Journal of Economics* 126: 2117-2123.

Barreca A, Lindo J, Waddell G. 2011. Heaping-induced bias in regression discontinuity designs. NBER Working Paper 17408.

Becker GS, Murphy KM. 1988. A theory of rational addition. *The Journal of Political Economy* 96: 675-700.

Beckett M, Da Vanzo J, Sastry N, Panis C, Peterson C. 2001. The quality of retrospective data: an examination of long-term recall in a developing country. *The Journal of Human Resources* 36: 593-625.

Bowers N, Horvath FW. 1984. Keeping time: an analysis of errors in the measurement of unemployment duration. *Journal of Business and Economic Statistics* 2: 140-149.

Chaloupka FJ, Warner KE. 2000. The economics of smoking. In *Handbook of Health Economics*, Culyer AJ, Newhouse JP (eds). Elsevier: Amsterdam.

Cosslett SR. 1993. Estimation from endogenously stratified samples. In *Handbook of Statistics*, Maddala GS, Rao CR, Vinod HD (eds). Elsevier: Amsterdam.

DeCicca P, Kenkel D, Mathios A. 2008. Cigarette taxes and the transition from youth to adult smoking: smoking initiation, cessation, and participation. *Journal of Health Economics* 27: 904-917.

Forster M, Jones AM. 2001. The role of tobacco taxes in starting and quitting smoking: duration analysis of British data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 164: 517-547.

Forster M, Smith SD. 2011. Surviving slavery: mortality at Mesopotamia, a Jamaican sugar estate, 1762-1832. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 174: 907-929.

Hausman JA. 2001. Mismeasured variables in econometric analysis: problems from the right and problems from the left. *Journal of Economic Perspectives* 15: 57-67.

Hausman JA, Abrevaya J, Scott-Morton FM. 1998. Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics* 87: 239-269.

Heitjan DF, Rubin DB. 1990. Inference from coarse data via multiple imputation with application to age heaping. *Journal of the American Statistical Association* 85: 304-314.

Hujer R, Maurer KO, Wellner M. 1997. Estimating the effect of training on unemployment duration in West Germany – a discrete hazard-rate model with instrumental variables. Johann Wolfgang Goethe-University Working Paper.

Hujer R, Schneider H. 1989. The analysis of labor market mobility using panel data. *European Economic Review* 33: 530-536.

Jenkins SP. 1995. Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics* 57: 129-136.

Jones AM. 1999. Adjustment costs, withdrawal effects, and cigarette addiction. *Journal of Health Economics* 18: 125-137.

Keane MP, Sauer RM. 2009. Classification error in dynamic discrete choice models: implications for female labor supply behavior. *Econometrica* 77: 975-991.

Kenkel DS, Lillard DR, Mathios AD. 2003. Smoke or fog? the usefulness of retrospectively reported information about smoking. *Addiction* 98: 1307-1313.

Kenkel DS, Lillard DR, Mathios AD. 2004. Accounting for misclassification error in retrospective smoking data. Health Economics 13: 1031-1044.

Kraus F, Steiner V. 1996. Modeling heaping effects in unemployment duration models – with an application to retrospective event data in the German socio-economic panel. ZEW Discussion Paper 95-09.

Krosnick JA. 1999. Survey research. *Annual Review of Psychology* 50: 537-567.

Langer EJ. 1992. Matters of mind: mindfulness/mindlessness in perspective. *Consciousness and Cognition* 1: 289-305.

Last JM. 2001. *A Dictionary of Epidemiology*. Oxford University Press, NY.

Machlin SR, Kleinman JC, Madans JH. 1989. Validity of mortality analysis based on retrospective smoking information. *Statistics in Medicine* 8: 997-1009.

Manski CF, Molinari F. 2010. Rounding probabilistic expectations in surveys. *Journal of Business and Economic Statistics* 28: 219-231.

Margolis J, Hockenberry J, Grossman M, Chou SY. 2014. Moral hazard and less invasive medical treatment for coronary artery disease: the case of cigarette smoking. NBER Working Paper 20373.

McLain AC, Sundaram R, Thoma M, Buck Louis GM. 2014. Semiparametric modeling of grouped current duration data with preferential reporting. *Statistics in Medicine* 33: 3961-3972.

Meyer B, Mittag N. 2014. Misclassification in binary choice models. NBER Working Paper 20509.

Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. 1994. The validity of self-reported smoking: a review and meta-analysis. *American Journal of Public Health* 84: 1086-1093.

Poterba JM, Summers LH. 1995. Unemployment benefits and labor market transitions: a multinomial logit model with errors in classification. *Review of Economics and Statistics* 77: 207-216.

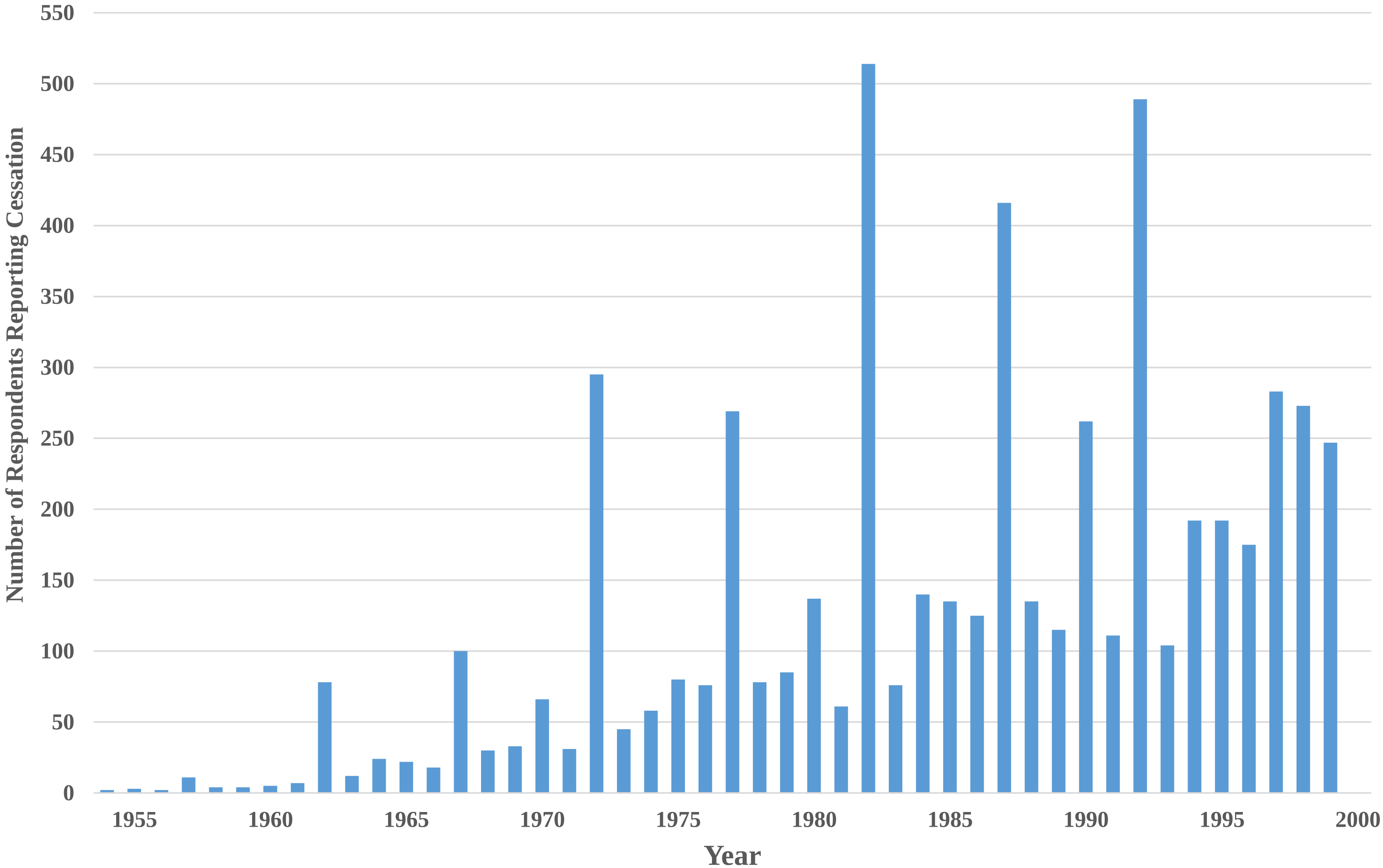
Ridout MS, Morgan BJT. 1991. Modelling digit preference in fecundability studies. *Biometrics* 47: 1423-1433.

Rubin DB. 1987. Multiple Imputation for Nonresponse in Survey. Wiley, NY.

Suranovic SM, Goldfarb RS, Leonard TC. 1999. An economic theory of cigarette addiction. *Journal of Health Economics* 18: 1-29.

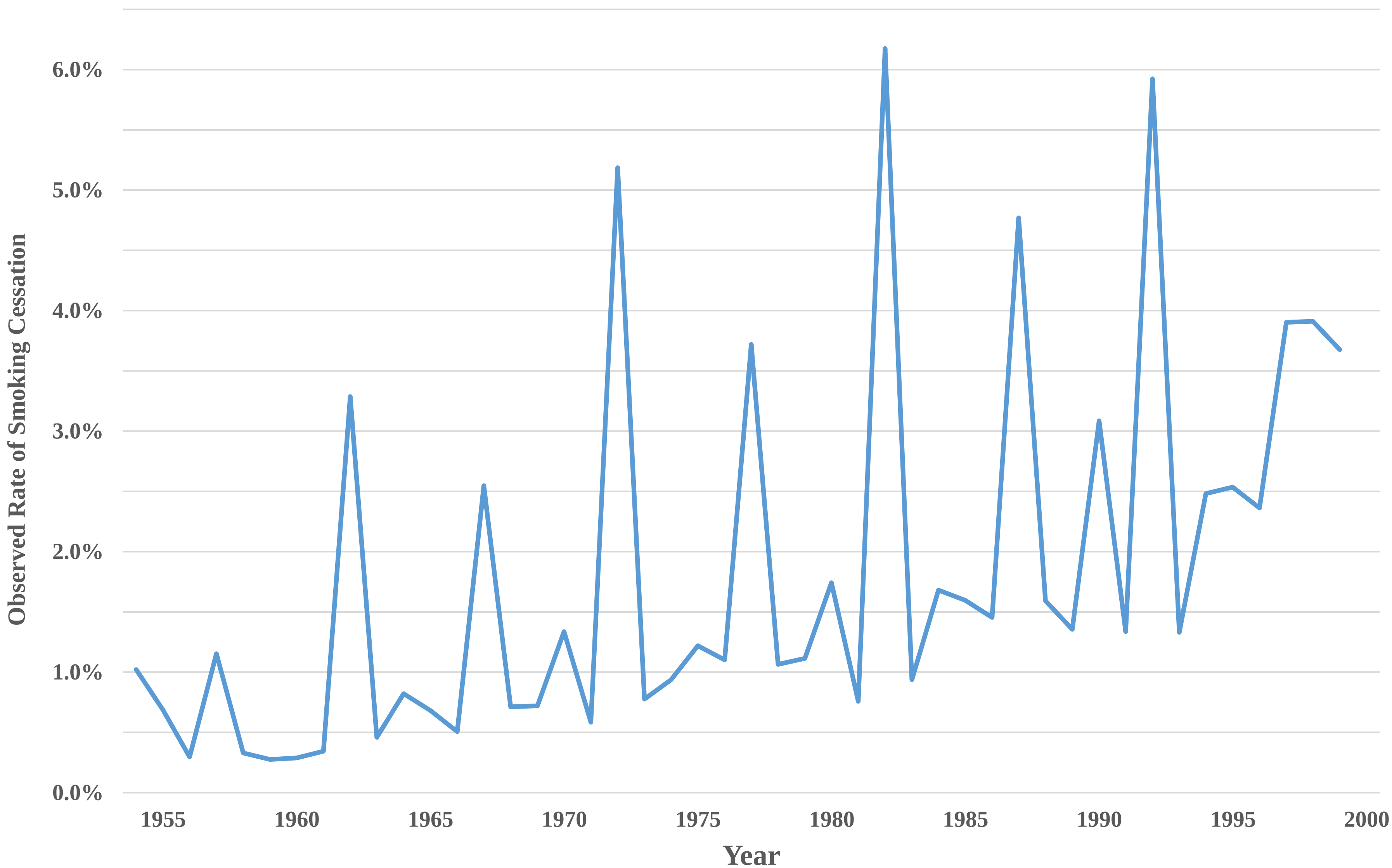
Torelli N, Trivellato U. 1993. Modelling inaccuracies in job-search duration data. *Journal of Econometrics* 59: 187-211.

Figure 1. Distribution of Recalled Year of Cessation among Former Smokers, CPS TUS 2002



Source. Tobacco Use Supplement to the 2002 Current Population Survey.

Figure 2. Observed Rate of Smoking Cessation, CPS TUS 2002



Source. Tobacco Use Supplement to the 2002 Current Population Survey.

Table 1. Coefficient Estimates from Simulated Smoking Hazard Data with Respondent Recall of "*How Many Years Ago Did You Quit Smoking?*" Randomly Heaped to *5* s and *10* s

Covariate	True	Simulation Results[1]	Method[2]						
			Probit on True Data	**Naïve Probit**	**Linear Probability Model**	**Probit with Heaping Indicator**	**Probit on Decimated Sample**	**Misclass-ification MLE Model**	**Probit after Multiple Imputation[3]**
Price of Cigarettes	**0.2000**	Mean	0.2001	0.1913	0.0377	0.1906	0.1822	0.2085	0.1761
		(Standard Deviation)	(0.0087)	(0.0087)	(0.0017)	(0.0088)	(0.0109)	(0.0101)	(0.0084)
		[% Diff. from True]	[0.03 %]	[-4.34 %]	[-81.15 %]	[-4.69 %]	[-8.88 %]	[4.27 %]	[-11.93 %]
Age	**0.0200**	Mean	0.0199	0.0198	0.0034	0.0196	0.0184	0.0210	0.0221
		(Standard Deviation)	(0.0012)	(0.0012)	(0.0002)	(0.0012)	(0.0015)	(0.0014)	(0.0012)
		[% Diff. from True]	[-0.25 %]	[-1.15 %]	[-83.11 %]	[-1.80 %]	[-7.78 %]	[5.14 %]	[10.57 %]
Female	**0.0300**	Mean	0.0301	0.0300	0.0051	0.0313	0.0374	0.0356	0.0275
		(Standard Deviation)	(0.0124)	(0.0123)	(0.0022)	(0.0126)	(0.0163)	(0.0142)	(0.0125)
		[% Diff. from True]	[0.26 %]	[-0.09 %]	[-83.00 %]	[4.42 %]	[24.68 %]	[18.56 %]	[-8.20 %]
Years of Education	**0.0800**	Mean	0.0801	0.0772	0.0124	0.0779	0.0782	0.0856	0.0774
		(Standard Deviation)	(0.0029)	(0.0028)	(0.0005)	(0.0029)	(0.0035)	(0.0033)	(0.0029)
		[% Diff. from True]	[0.15 %]	[-3.56 %]	[-84.47 %]	[-2.61 %]	[-2.21 %]	[6.95 %]	[-3.24 %]
Household Income[4]	**0.0080**	Mean	0.0080	0.0080	0.0015	0.0081	0.0086	0.0091	0.0080
		(Standard Deviation)	(0.0003)	(0.0003)	(0.0000)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
		[% Diff. from True]	[0.07 %]	[-0.42 %]	[-81.68 %]	[1.22 %]	[7.10 %]	[13.69 %]	[-0.47 %]
Year[5]	**0.0200**	Mean	0.0200	0.0188	0.0019	0.0190	0.0283	0.0214	0.0199
		(Standard Deviation)	(0.0027)	(0.0027)	(0.0004)	(0.0027)	(0.0033)	(0.0031)	(0.0026)
		[% Diff. from True]	[0.06 %]	[-6.20 %]	[-90.27 %]	[-4.98 %]	[41.32 %]	[7.15 %]	[-0.53 %]
Year, Squared	**-0.0003**	Mean	-0.0003	-0.0003	0.0000	-0.0003	-0.0005	-0.0004	-0.0003
		(Standard Deviation)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
		[% Diff. from True]	[-0.45 %]	[-8.84 %]	[-91.91 %]	[-7.48 %]	[82.55 %]	[27.06 %]	[-3.36 %]
Constant	**-3.8000**	Mean	-3.8018	-3.7248	-0.2941	-3.9010	-3.7460	-3.9003	-3.7644
		(Standard Deviation)	(0.0520)	(0.0523)	(0.0070)	(0.0536)	(0.0658)	(0.0599)	(0.0534)
		[% Diff. from True]	[0.05 %]	[-1.98 %]	[-92.26 %]	[2.66 %]	[-1.42 %]	[2.64 %]	[-0.94 %]

Notes. 1. Results are from 250 simulations with 10,000 respondents. 2. See text for further description of the methods and Monte Carlo design. 3. Analysis uses 5 imputations of respondent recall. 4. Household income is measured in $10,000s. 5. Starting with 1954, years are recoded for analysis to begin at 1.

Table 2. Additional Coefficient and Parameter Estimates from Simulated Smoking Hazard Data with Respondent Recall of "*How Many Years Ago Did You Quit Smoking?* " Randomly Heaped to *5* s and *10* s

Covariate	Simulation Results[1]	Method[2]: Probit with Heaping Indicator	Method[2]: Misclass-ification MLE Model
Heap Year Indicator	Mean	0.5918	
	(Standard Deviation)	(0.0134)	
α_{0H}: Year[3] (λ_1)	Mean		0.0528
	(Standard Deviation)		(0.0058)
α_{0H}: Constant (λ_0)	Mean		-3.9688
	(Standard Deviation)		(0.1849)
α_{1N}: 2 Years Before Heap (μ_2)	Mean		0.4279
	(Standard Deviation)		(0.0182)
α_{1N}: 1 Year Before Heap (μ_1)	Mean		0.4585
	(Standard Deviation)		(0.0182)

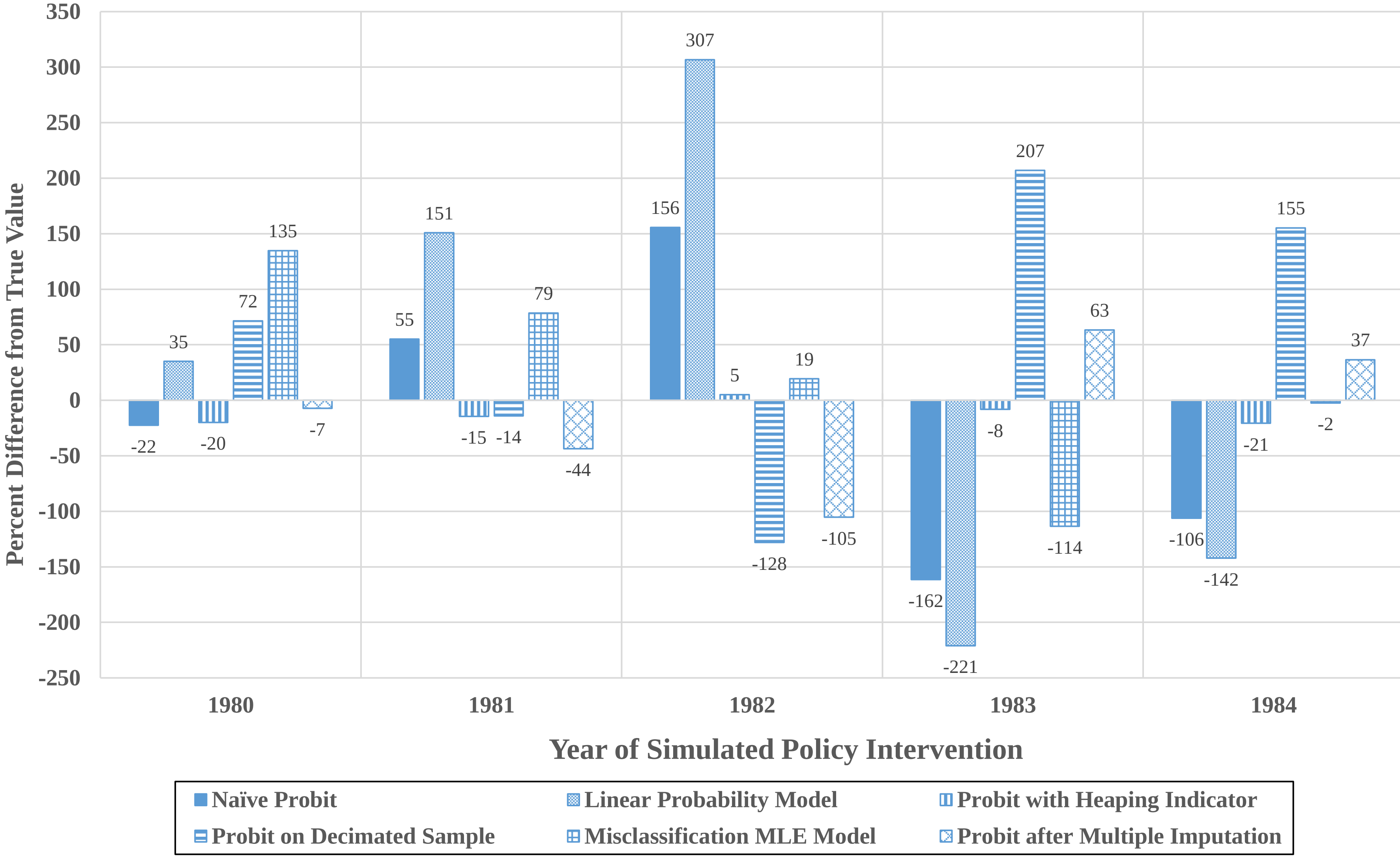
Notes. 1. Results are from 250 simulations with 10,000 respondents. 2. See text for further description of the methods and Monte Carlo design. 3. Starting with 1954, years are recoded for analysis to begin at 1.

Table 3. Marginal Effect Estimates from the First Year of Observation of Each Respondent in the Simulated Smoking Hazard Data with Respondent Recall of "*How Many Years Ago Did You Quit Smoking?*" Randomly Heaped to *5* s and *10* s

Covariate	True[1]	Simulation Results[2]	Method[3]: Probit on True Data	Naïve Probit	Linear Probability Model	Probit with Heaping Indicator	Probit on Decimated Sample	Misclass-ification MLE Model	Probit after Multiple Imputation[4]
Price of Cigarettes	**0.0247**	Mean	0.0247	0.0239	0.0377	0.0232	0.0244	0.0273	0.0218
		(Standard Deviation)	(0.0010)	(0.0010)	(0.0017)	(0.0010)	(0.0014)	(0.0012)	(0.0010)
		[% Diff. from True]	[-0.22 %]	[-3.23 %]	[52.56 %]	[-6.09 %]	[-1.39 %]	[10.57 %]	[-11.72 %]
Age	**0.0025**	Mean	0.0025	0.0025	0.0034	0.0024	0.0025	0.0028	0.0027
		(Standard Deviation)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)
		[% Diff. from True]	[-0.50 %]	[-0.01 %]	[36.73 %]	[-3.24 %]	[-0.22 %]	[11.50 %]	[10.84 %]
Female	**0.0037**	Mean	0.0037	0.0037	0.0051	0.0038	0.0050	0.0047	0.0034
		(Standard Deviation)	(0.0015)	(0.0016)	(0.0022)	(0.0015)	(0.0022)	(0.0019)	(0.0016)
		[% Diff. from True]	[0.07 %]	[1.15 %]	[37.55 %]	[2.96 %]	[35.06 %]	[25.81 %]	[-7.90 %]
Years of Education	**0.0099**	Mean	0.0099	0.0096	0.0124	0.0095	0.0105	0.0112	0.0096
		(Standard Deviation)	(0.0003)	(0.0003)	(0.0005)	(0.0004)	(0.0005)	(0.0004)	(0.0003)
		[% Diff. from True]	[-0.09 %]	[-2.43 %]	[25.67 %]	[-4.03 %]	[5.85 %]	[13.43 %]	[-2.99 %]
Household Income[5]	**0.0010**	Mean	0.0010	0.0010	0.0015	0.0010	0.0011	0.0012	0.0010
		(Standard Deviation)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
		[% Diff. from True]	[-0.16 %]	[0.76 %]	[48.27 %]	[-0.24 %]	[15.94 %]	[20.60 %]	[-0.21 %]

Notes. 1. True value determined by calculating marginal effects for the first year of observation for all respondents in the 2002 CPS TUS sample (used in the bootstrap sampling) using the covariate coefficients from the simulation. 2. Results are from 250 simulations with 10,000 respondents. 3. See text for further description of the methods and Monte Carlo design. 4. Analysis uses 5 imputations of respondent recall. 5. Household income is measured in $10,000s.

Figure 3. Bias in the Estimated Marginal Effects of The Simulated Policy Intervention, 1980–1984



Notes. 1. True value determined by calculating marginal effects for the first year of observation for all respondents in the 2002 CPS TUS sample (used in the bootstrap sampling) using the covariate coefficients from the simulation. 2. Results are from 250 simulations with 10,000 respondents. 3. See text for further description of the methods and Monte Carlo design. 4. Analysis uses 5 imputations of respondent recall in the Multiple Imputation approach.

Table A1. Summary Statistics for the Sample from the Tobacco Use Supplement to the 2002 Current Population Survey [1]

Respondent Characteristics in 1999 [2]	Sample Mean (Standard Deviation)	Observation-Level Characteristics	Sample Mean (Standard Deviation)
Age	44.7414 (10.1430)	**Real Price of Cigarettes**	1.9437 (0.4012)
Age at Smoking Initiation	18.0081 (4.5337)	**Age**	30.5243 (10.2271)
Age at Smoking Cessation (Among Quitters)	34.0931 (10.6392)	**Female**	0.4838 (0.4997)
Observed Duration of Smoking (Years, Full Sample)	23.5260 (11.6739)	**Years of Education**	12.5069 (2.4761)
Observed Duration of Smoking (Years, Quitters Only)	16.9415 (10.1870)	**Household Income**	45.5064 (26.0034)
Year of Smoking Initiation	1972.2667 (9.9574)	**Year** [4]	29.2423 (10.5744)
Year of Smoking Cessation (Among Quitters)	1985.5217 (9.6003)		
Female	0.4890 (0.4999)		
Total Years of Education	13.0754 (2.5142)		
Household Income in 2002 [3]	47.9431 (26.1483)		

Notes. 1. Tobacco Use Supplement to the 2002 Current Population Survey; 12,093 respondents provide 259,631 total observations. 2. Summary statistics are presented for 1999, the final year prior to right-censoring in the analysis. 3. Household income is measured in $10,000s for the survey year, 2002. 4. Starting with 1954, years are recoded for analysis to begin at 1.

Table A2. Coefficient and Average Marginal Effect Estimates from Discrete-Time Hazard Analysis of Smoking Cessation Reconstructed from the Tobacco Use Supplement to the 2002 Current Population Survey

Covariate[1]		**Coefficient**[2]	**Average Marginal Effect**[3]
Price of Cigarettes	Estimate	0.1009	0.0051
	(Standard Error)	(0.0208)	(0.0010)
Age	Estimate	0.0103	0.0005
	(Standard Error)	(0.0007)	(0.0000)
Female	Estimate	0.0212	0.0011
	(Standard Error)	(0.0113)	(0.0006)
Years of Education	Estimate	0.0423	0.0021
	(Standard Error)	(0.0025)	(0.0001)
Household Income[4]	Estimate	0.0049	0.0002
	(Standard Error)	(0.0002)	(0.0000)
Year[5]	Estimate	0.0161	
	(Standard Error)	(0.0034)	
Year, Squared	Estimate	-0.0002	
	(Standard Error)	(0.0001)	
Constant	Estimate	-3.6366	
	(Standard Error)	(0.0705)	

Notes. 1. Tobacco Use Supplement to the Current Population Survey; 12,093 respondents. 2. All estimates are statistically significant at the $\alpha = 1\%$ level of significance except *Female* , which is statistically significant at $\alpha = 10\%$. 3. Average marginal effects reported for covariates of interest only. 4. Household income is measured in $10,000s. 5. Starting with 1954, years are recoded for analysis to begin at 1.

Table A3. Marginal Effect Estimates from the First Year of Observation of Each Respondent in the Simulated Smoking Hazard Data with Respondent Recall of "*How Many Years Ago Did You Quit Smoking?*" Randomly Heaped More to *10* s than *5* s

			Method[3]						
Covariate	**True**[1]	**Simulation Results**[2]	**Probit on True Data**	**Naïve Probit**	**Linear Probability Model**	**Probit with Heaping Indicator**	**Probit on Decimated Sample**	**Misclass-ification MLE Model**	**Probit after Multiple Imputation**[4]
Price of Cigarettes	**0.0247**	Mean	0.0247	0.0236	0.0370	0.0219	0.0237	0.0253	0.0210
		(Standard Deviation)	(0.0010)	(0.0010)	(0.0017)	(0.0010)	(0.0016)	(0.0013)	(0.0010)
		[% Diff. from True]	[-0.22 %]	[-4.69 %]	[49.58 %]	[-11.54 %]	[-4.13 %]	[2.33 %]	[-14.84 %]
Age	**0.0025**	Mean	0.0025	0.0025	0.0034	0.0023	0.0025	0.0026	0.0028
		(Standard Deviation)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)
		[% Diff. from True]	[-0.50 %]	[0.29 %]	[36.85 %]	[-6.55 %]	[1.68 %]	[5.37 %]	[13.52 %]
Female	**0.0037**	Mean	0.0037	0.0038	0.0051	0.0038	0.0060	0.0049	0.0034
		(Standard Deviation)	(0.0015)	(0.0016)	(0.0022)	(0.0015)	(0.0026)	(0.0020)	(0.0016)
		[% Diff. from True]	[0.07 %]	[1.53 %]	[37.59 %]	[3.75 %]	[60.79 %]	[31.44 %]	[-9.53 %]
Years of Education	**0.0099**	Mean	0.0099	0.0095	0.0122	0.0091	0.0109	0.0108	0.0095
		(Standard Deviation)	(0.0003)	(0.0004)	(0.0005)	(0.0004)	(0.0006)	(0.0005)	(0.0004)
		[% Diff. from True]	[-0.09 %]	[-3.52 %]	[23.90 %]	[-7.60 %]	[10.01 %]	[9.23 %]	[-4.31 %]
Household Income[5]	**0.0010**	Mean	0.0010	0.0010	0.0015	0.0010	0.0013	0.0012	0.0010
		(Standard Deviation)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)
		[% Diff. from True]	[-0.16 %]	[1.06 %]	[48.03 %]	[-1.39 %]	[29.99 %]	[22.98 %]	[-0.15 %]

Notes. 1. True value determined by calculating marginal effects for the first year of observation for all respondents in the 2002 CPS TUS sample (used in the bootstrap sampling) using the covariate coefficients from the simulation. 2. Results are from 250 simulations with 10,000 respondents. 3. See text for further description of the methods and Monte Carlo design. 4. Analysis uses 5 imputations of respondent recall. 5. Household income is measured in $10,000s.

Table A4. Marginal Effect Estimates from the First Year of Observation of Each Respondent in the Simulated Smoking Hazard Data with Respondent Recall of "*How Many Years Ago Did You Quit Smoking?*" Randomly Heaped to *5* s, *10* s, and *12* , *22* ,… Representing Calendar Decades

Covariate	True[1]	Simulation Results[2]	Method[3]: Probit on True Data	Naïve Probit	Linear Probability Model	Probit with Heaping Indicator	Probit on Decimated Sample	Misclass-ification MLE Model	Probit after Multiple Imputation[4]
Price of Cigarettes	**0.0247**	Mean	0.0247	0.0241	0.0380	0.0237	0.0244	0.0267	0.0222
		(Standard Deviation)	(0.0010)	(0.0010)	(0.0017)	(0.0010)	(0.0013)	(0.0012)	(0.0010)
		[% Diff. from True]	[-0.22 %]	[-2.54 %]	[53.59 %]	[-4.27 %]	[-1.15 %]	[7.87 %]	[-9.97 %]
Age	**0.0025**	Mean	0.0025	0.0025	0.0034	0.0024	0.0024	0.0027	0.0027
		(Standard Deviation)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0002)
		[% Diff. from True]	[-0.50 %]	[-0.16 %]	[36.41 %]	[-2.34 %]	[-1.19 %]	[8.42 %]	[10.19 %]
Female	**0.0037**	Mean	0.0037	0.0038	0.0051	0.0038	0.0047	0.0044	0.0035
		(Standard Deviation)	(0.0015)	(0.0016)	(0.0022)	(0.0015)	(0.0021)	(0.0018)	(0.0016)
		[% Diff. from True]	[0.07 %]	[1.40 %]	[37.83 %]	[2.90 %]	[27.79 %]	[19.78 %]	[-5.94 %]
Years of Education	**0.0099**	Mean	0.0099	0.0097	0.0125	0.0096	0.0104	0.0109	0.0098
		(Standard Deviation)	(0.0003)	(0.0004)	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
		[% Diff. from True]	[-0.09 %]	[-1.60 %]	[26.70 %]	[-2.48 %]	[4.91 %]	[10.36 %]	[-1.35 %]
Household Income[5]	**0.0010**	Mean	0.0010	0.0010	0.0015	0.0010	0.0011	0.0011	0.0010
		(Standard Deviation)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
		[% Diff. from True]	[-0.16 %]	[0.94 %]	[48.59 %]	[0.40 %]	[12.17 %]	[15.64 %]	[0.94 %]

Notes. 1. True value determined by calculating marginal effects for the first year of observation for all respondents in the 2002 CPS TUS sample (used in the bootstrap sampling) using the covariate coefficients from the simulation. 2. Results are from 250 simulations with 10,000 respondents. 3. See text for further description of the methods and Monte Carlo design. 4. Analysis uses 5 imputations of respondent recall. 5. Household income is measured in $10,000s.