JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **21**: 489–519 (2006)
Published online 8 March 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.851

ESTIMATING THE EFFECT OF SMOKING ON BIRTH OUTCOMES USING A MATCHED PANEL DATA APPROACH

JASON ABREVAYA*
Department of Economics, Purdue University, West Lafayette, IN 47907-2056, USA

SUMMARY

Estimating the casual effect of smoking on birth outcomes is difficult since omitted (unobserved) variables are likely to be correlated with a mother's decision to smoke. While some previous work has dealt with this endogeneity problem by using instrumental variables, this paper instead attempts to estimate the smoking effect from panel data (i.e., data on mothers with multiple births). Panel data sets are constructed with matching algorithms applied to federal natality data. The fixed effects regressions, which control for individual heterogeneity, yield significantly different results from ordinary least squares and previous instrumental variable approaches. The potential inconsistency caused by 'false matches' and other violations of the fixed effects strict exogeneity assumption are considered. Copyright © 2006 John Wiley & Sons, Ltd.

1. INTRODUCTION

A 2001 report by the Surgeon General summarizes the relationship between smoking during pregnancy and birth outcomes as follows: 'Infants born to women who smoke during pregnancy have a lower average birthweight and are more likely to be small for gestational age than infants born to women who do not smoke. Low birthweight is associated with increased risk for neonatal, perinatal, and infant morbidity and mortality. The longer the mother smokes during pregnancy, the greater the effect on the infant's birthweight... [S]moking is also associated with a modest increase in risk for preterm delivery.' (*Women and Smoking: A Report of the Surgeon General*, Centers for Disease Control and Prevention, 2001).

The issue of low birthweight has received a great deal of attention in the economics literature due to the direct medical costs and long-term costs associated with low birthweight. While low-birthweight (LBW) infants (defined as weighing less than 2500 g at birth) account for less than 10% of births in the United States, Lewit *et al.* (1995) estimate that LBW infants account for more than one-third of health care costs during the first year of life. Joyce (1999) estimates that the marginal cost of hospital newborn care is \$6–10/g, with higher marginal costs at lower birthweights. Several studies have shown that lower birthweight is associated with higher infant mortality rates (see, for instance, the references in Mathews (2001)). Schoendorf and Kiely (1992) link maternal smoking to higher rates of sudden infant death syndrome.

Recent research has also shown that, in addition to the short-term costs, low birthweight can have substantial long-term effects as well. Hack *et al.* (1995) find that low-birthweight babies have developmental problems in cognition, attention and neuromotor functioning that persist

* Correspondence to: Jason Abrevaya, Department of Economics, Purdue University, 403 W. State Street, West Lafayette, IN 47907-2056, USA. E-mail: abrevaya@purdue.edu
Contract/grant sponsor: John & Mary Willis Young Faculty Scholar Fund, Purdue University.
Contract/grant sponsor: Faculty research Fund, University of Chicago GSB.

Received 3 February 2004
Revised 2 November 2004

until adolescence. Corman (1995) and Corman and Chaikind (1998) find that low-birthweight children are more likely to delay entry into kindergarten, repeat a grade in school and attend special education classes. Currie and Hyson (1999) also find an effect of low birthweight on eventual labour market outcomes (in addition to the effects on health status and education that they find). With these long-term effects in mind, it is possible that low birthweight even exacerbates socioeconomic inequality, since less privileged mothers are those most likely to have low-birthweight babies.

The true benefits of any public health initiative focused on reducing smoking during pregnancy depend crucially upon the causal effect of smoking on birth outcomes. While it is relatively easy to establish a negative relationship between maternal smoking and birth outcomes, it is far more difficult to accurately estimate the causal effect of maternal smoking on a birth outcome like birthweight. The problem is that omitted variables which also affect birthweight are likely to be correlated with a mother's decision to smoke. To fix ideas, consider the following model describing the birthweight y associated with a given birth:

$$y = x'\beta + \gamma s + \varepsilon \tag{1}$$

where x is a vector of explanatory variables, s is a smoking indicator variable (equal to one if the mother smokes during pregnancy and zero if not) and ε is an error disturbance. An ordinary least squares (OLS) regression of y on x and s is undoubtedly subject to an omitted variables bias. One would expect, for instance, that mothers who smoke during pregnancy (births with $s = 1$) are more likely to adopt other behaviours (drinking, poor nutritional intake, etc.) that could have a negative impact on birthweight. That is, one would expect that s and ε are negatively correlated, meaning that the OLS estimator $\hat{\gamma}$ likely overestimates the causal effect of s. In practice, an OLS regression of the model in (1) yields an estimate of $\hat{\gamma}$ around -250 g. If s and ε are negatively correlated, this estimate is larger in magnitude than the causal effect since s is also proxying for unobservables that adversely affect y.

Previous attempts at estimating the causal effect of smoking on birth outcomes have used an instrumental variables (IV) strategy, instrumenting for s in equation (1).[1] Permutt and Hebel (1989), for instance, report results from a study in which there was randomized 'intervention' (in the form of counselling) for a treatment group chosen from a group of smoking women: 43% of the treatment subjects stopped smoking, as opposed to 20% of the control subjects (i.e., smoking women not offered counselling). The associated causal effect of smoking on birthweights is a reduction of 400 g (with a standard error of 177 g). Since only 935 women were included in the study, the resulting estimate of the causal effect is rather imprecise.

Utilizing a much larger data set (over 10 million births from federal natality data), Evans and Ringel (1999) use state cigarette taxes as an instrument for smoking during pregnancy. In fact, their study is the first to provide evidence that higher cigarette taxes lead to a reduction in smoking during pregnancy. The second-stage estimates of the effect of smoking on birthweight indicate that smoking lowers birthweight by between 350 g and 600 g. Though these estimated effects are larger in magnitude than the OLS estimates (230–250 g effects in their study), the difference between the IV estimates and OLS estimates is not statistically significant. Despite the large amount of data used, the identification of the smoking effect in Evans and Ringel (1999) comes

[1] Other studies have used instrumental variables in order to estimate the effect of prenatal care on birth outcomes. These include Currie and Gruber (1996), Evans and Lien (2005), Levinson and Ullman (1998), Reichman and Florio (1996). For a critical review of medical attempts at estimating the causal effect of smoking, see Walsh (1994).

from changes in state cigarette tax rates, which are relatively infrequent during the period of their study (1989–1992).

In a related study, Lien and Evans (2001) use a more recent sample from the federal natality data (1990–1997) and focus on the effect of large cigarette tax changes that occurred in four states (Arizona, Illinois, Michigan and Massachusetts). By comparing outcomes in these ('treatment') states to those in similar ('control') states, the authors estimate that smoking reduces birthweight by 189 g, which is statistically indistinguishable from the OLS estimates of smoking on birthweight. As Lien and Evans (2001) stress, this estimate applies to the group of women who would be likely to quit smoking as a result of a tax hike. The authors comment that 'these results suggest that the omitted variable bias in single equation models where maternal smoking is treated as exogenous is not severe'.

Rather than using instrumental variables estimation, this paper instead attempts to estimate the effect of smoking on birth outcomes from panel data (i.e., data on mothers with multiple births). Panel data allows the identification of the smoking effect from women who change their smoking behaviour from one pregnancy to another. In contrast, the identification from IV estimation relies on a change in smoking behaviour that is caused by the specific instrument (e.g., cigarette taxation) being examined. Like Evans and Ringel (1999) and Lien and Evans (2001), we utilize federal natality data for our analysis. This study considers data on birth outcomes between 1990 and 1998.[2] The main obstacle to constructing a panel data set from the federal natality data is that there are no fields that can be used to uniquely identify a mother (e.g., social security number). As such, a matching strategy is employed in order to isolate individual mothers for whom multiple births can be identified during the time period.[3] The number of observations in the resulting matched panel data is extremely large, even though the constructed panels contain fewer than 1% of the observations from the original federal natality data.

As far as we know, the only other study to use panel data and report estimates of the effects of smoking during pregnancy is Rosenzweig and Wolpin (1991). Their panel is constructed from the National Longitudinal Survey of Youth (NLSY). Rosenzweig and Wolpin (1991) find that the estimated effect of smoking is lower in magnitude than the OLS estimate, but their rather small sample (3384 births) results in t-statistics similar in magnitude to the aforementioned IV studies.

Figure 1 summarizes the estimated effect of smoking on birthweight from the previous approaches discussed above. For each study, the point estimate is shown (bold circle) along with 95% confidence intervals (lines drawn at plus/minus two standard errors).[4] As a point of comparison, OLS estimates of smoking's negative effect on birthweight are generally in the 200–250 g range. Interestingly, the IV approaches of Permutt and Hebel (1989) and Evans and Ringel (1999) actually yield point estimates that are larger than the OLS estimates, contrary to what would be expected from the basic omitted variables story.[5]

[2] Smoking information was not provided in the federal natality data prior to 1989, and 1998 was the last year of data available when this project was started.

[3] In independent work, Currie and Moretti (2002) use a similar matching strategy in order to estimate the effect of maternal education on birth outcomes from panel data. Due to the time period considered in their study and the lack of smoking data prior to 1989, smoking variables are not considered in their panel data estimation results.

[4] For the Evans and Ringel (1999) estimate, we cite their results from Model (2) in Table III of their paper. For Lien and Evans (2001), we cite the aggregate result from pooling the four states. For Rosenzweig and Wolpin (1991), we cite their estimate for women who smoked at least one pack per day.

[5] As Evans and Ringel (1999) point out, the estimated impact of smoking may be larger in magnitude since the IV strategy identifies the effect from mothers who quit smoking in response to cigarette taxes. If the actual effect of smoking for these mothers is larger, the estimated causal effect could be larger in magnitude than the OLS estimate.

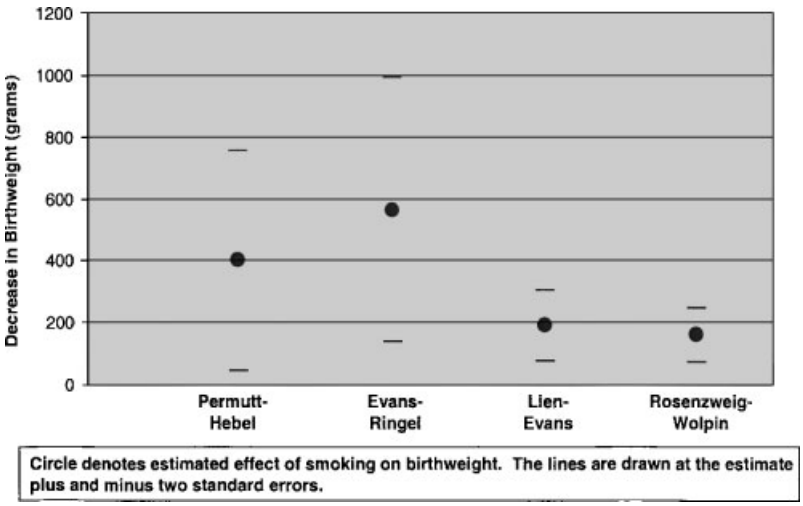

Figure 1. Previous estimates of smoking's effect on birthweight

The remainder of the paper is outlined as follows. Section 2 describes the data and the matching strategies used to construct panel data sets for analysis. Section 3 considers the inconsistency caused by incorrect matching. The direction of this inconsistency is the same as the omitted variables bias, and a simple model suggests that the mismatched observations, 'false matches', have a disproportionate effect on the inconsistency. Section 4 reports the main empirical results. First, fixed effects estimates (based upon the matched panels of Section 2) are reported. Second, using a proxy for correct matching that is available in the earlier part of the sample, the effect of mismatching is empirically investigated. Third, an augmented model specification is used to allow for the possibility of heterogeneous smoking effects. Section 5 considers several possible violations of the strict exogeneity assumption made for fixed effects estimation. These potential violations include feedback effects (smoking influenced by a prior birth outcome), correlated maternal behaviour (smoking changes correlated with unobservable behavioural changes) and misclassification of smoking status. Section 6 concludes.

2. DATA AND MATCHING STRATEGY

2.1. Federal Natality Data

The data used for this study come from the Natality Data Sets (released by the National Center for Health Statistics (NCHS)) from 1990 to 1998. The natality data are based upon birth records from every live birth that occurs in the United States and contain information on birth outcomes, maternal prenatal behaviour (including smoking behaviour for nearly all states) and demographic attributes. During the time period analysed, there are approximately 4 million births per year in the natality data.

Unfortunately, the federal natality data does have limitations. There are no unique identifiers (e.g., social security number) for mothers, which makes construction of a panel data set non-trivial. Other (non-unique) identifiers that would make matching far easier (such as mother's name and

mother's date of birth) are also absent from the data for confidentiality reasons. Although the demographic information includes age, education level, race, marital status and state of residency, there is no direct information on income and only imprecise measures of geographic location. Specific city and/or county location are given only if the population is greater than 100 000; if not, city and/or county are listed as 'other'.

The questions related to smoking during pregnancy have some drawbacks.[6] First, four states (California, Indiana, New York and South Dakota) either did not ask about smoking status or did not ask about it in a form compatible with NCHS standards (Mathews, 2001). As such, births in these states are not considered. Second, the question about smoking status on the standard birth certificate is not very specific, as it asks only whether or not there was 'tobacco use during pregnancy' and, if so, the number of cigarettes smoked per day. The new standard birth certificate, which was introduced in 2002, asks about smoking behaviour in the three months prior to pregnancy and in each of the three trimesters of pregnancy.[7]

In addition to the matching requirements detailed below in Section 2.2, we consider only singleton births (i.e., no twins or other multiple births) and births to mothers whose race was classified as white or black. Furthermore, in order to deal with the possible selection issue that longer gestation births would be observed at the beginning of the data window (and shorter gestation births at the end of the data window), we proceed similarly to Evans and Ringel (1999) and consider only those births whose conception (based upon birth month and gestation estimate) occurred between August 1989 and January 1998.

2.2. Construction of Matched Panel Data Sets

The lack of mother-specific identifiers makes it difficult to match multiple births of the same mother. In order to ensure that our matching algorithms are somewhat reliable (i.e., result in 'true' matches), we first focus on a specific data item from the natality data—*mother's state of birth*. For every birth, there is an associated state-pair corresponding to the baby's state of birth and the mother's state of birth. Certain state-pairs (such as those with baby and mother born in the same state) are very common, whereas other state-pairs are not very common. As an example, there were 61 908 births in Alabama in 1990. Of these births, 44 652 were to mothers who were also born in Alabama; in contrast, far fewer births occur for mothers born in other states (from a minimum of 11 (Vermont) to a maximum of 2212 (Georgia)). To fully illustrate this point, Table XII (provided in the appendix) details the number of births in Alabama by mother's state of birth. For mothers born in Alabama and even neighbouring states like Florida and Georgia, it will be more difficult to accurately link their births over time. Recognizing this fact, attention is restricted to state-pairs

[6] Aside from cigarette use, the federal natality data does not provide reliable information about other forms of substance abuse. Although the standard birth certificate contains questions regarding alcohol use, it is well known that there is severe under-reporting of alcohol usage during pregnancy. Other types of drug use (e.g., marijuana, cocaine) are not recorded in the federal natality data.

[7] This more detailed data will allow researchers to better understand the effects of smoking at different times before and during pregnancy. These more specific questions have been asked in smaller surveys such as the National Pregnancy and Health Survey (NPHS) and the Pregnancy Risk Assessment Monitoring System (PRAMS). Using PRAMS data, Colman *et al*. (2003) consider the impact of cigarette taxes on smoking behaviour before, during and after pregnancy.

that have small birth counts. In particular, only state-pairs with fewer than 500 births in 1990 are included in the data.[8] This restriction leaves 1762 of the original 2209 state-pairs.[9]

For all births within the restricted subset of state-pairs, matching algorithms were used in an effort to link a mother's births across time. Three different algorithms, with successively more stringent matching requirements, were employed and are summarized in Table I. For each of the algorithms, matching criteria are used in order to link any given birth (observation) with other births that are possible matches. Two births that are linked together are then retained for the matched panel data set if the link connecting the two births is unique. Each of the three algorithms requires an exact match for mother's state of birth, mother's race and state/county/city of birth and consistency of mother's age and birth order. For birth order, the natality data gives the 'live birth order' and consistency with respect to this variable means that the birth order differs by one across two births. For mother's age, the natality data only gives information on the month of birth and the age of the mother in years (at the time of birth). Therefore, a link between two births is possible if the age and birth month value combinations are possible. As an example, a 27-year-old woman who gives birth in April 1992 could be either 29 or 30 years old in October 1994; in this case, the first birth could be linked with births in October 1994 occurring to either 29-year-old or 30-year-old women.

The three matching algorithms differ in their treatment of mother's education and mother's marital status. The first matching algorithm, used to construct 'Matched panel #1', allows for links between two births for which the difference in years of education is consistent with the timespan between the two births.[10] While this approach allows for the possibility that years of education increases between births, it also increases the likelihood of 'false matches' since increases in

Table I. Matching algorithms

	Matched panel #1	Matched panel #2	Matched panel #3
Exact match:	Mother's state of birth Mother's race State of birth County of birth City of birth	Mother's state of birth Mother's race State of birth County of birth City of birth Mother's education	Mother's state of birth Mother's race State of birth County of birth City of birth Mother's education Marital status If married, father's race
Consistency:	Birth order Mother's age Mother's education	Birth order Mother's age	Birth order Mother's age If married, father's age
# of observations	171 319	229 867	296 218
# of mothers	82 595	110 260	141 929
# of 2-birth mothers	76 466	100 913	129 569

[8] We experimented with other versions of this 'cutoff' rule and found no important differences in either sample composition or the resulting estimates reported in the next section.

[9] The number of original state-pairs is $47 \times 47 = 2209$ since: (i) four states were dropped due to lack of smoking information and (ii) the District of Columbia is included as a 'state' on the birth certificate.

[10] This algorithm is most similar to the approach of Currie and Moretti (2002), who do not restrict education to be constant across births since their focus is estimating the effect of additional education on birth outcomes. The primary differences between their approach and the approach in this paper are: (i) they consider only women aged 16–24 at first birth and 17–30 at second birth and (ii) we do an initial 'cut' based upon uncommon state pairs.

educational attainment are not very common after childbirth, especially for women over 18. As Currie and Moretti (2002) report, the National Longitudinal Survey of Youth (NLSY) indicates that only 15% of women aged 19–20 and 7% of women aged 21–24 report additional schooling between births.[11] To explicitly deal with this concern, the second matching algorithm ('Matched panel #2') requires an exact match for years of education (rather than consistency) across births. This treatment of education is the only difference between the algorithms for constructing 'Matched panel #1' and 'Matched panel #2'. Finally, the third matching algorithm ('Matched panel #3') also requires an exact match on a mother's marital status across births and, if the mother is married, an exact match on father's race and consistency of father's age (which is also reported in years in the data). The rationale behind these additions is similar to the rationale for requiring an exact match on education. Women tend to retain the same marital status (and, if married, the same husband) across births; as a result, the first two matching algorithms are more prone to 'false matches' by allowing marital status (or husband) to change between births.

It is important to consider alternative matching strategies for two reasons. First, it allows us to gauge whether or not estimation results are sensitive to the way in which the matched panel data set is constructed. In Section 4, results are reported for each of the three matched panel data sets to allow for a comparison. Second, the use of different matching algorithms emphasizes an inherent tradeoff involved in trying to match data across time based upon observables. The tradeoff is that stricter matching requirements generally result in fewer 'false matches', but the resulting matched panel focuses on a more specific subpopulation. For example, 'Matched panel #2' considers only mothers for whom years of education does not change from one birth to the next. As a result, fixed effects results based upon this panel will not be able to identify the possible marginal effects of education on birth outcomes.

Interestingly, in our application, the stricter matching requirements actually yield more observations in the resulting panels. The stricter matching requirements yield fewer match links among the birth observations, but since only unique matches are kept, the net effect is that more observations ultimately end up in the panels having the stricter matching requirements. The bottom of Table I reports the sample sizes (total number of birth observations) for the three matched panels. (The number of mothers and mothers with exactly two births is also reported.) 'Matched panel #3' (296 218 births) is nearly twice the size of 'Matched panel #1' (171 319 births). Although the resulting sample sizes are extremely large, note that they represent only about 0.5–0.8% of all births occurring in the United States between 1990 and 1998. A possible concern is that these subsamples may not be representative of the whole sample, an issue considered in detail in the next subsection.

2.3. Comparison with a Random Sample of Births

The construction of the matched panel data sets results in samples with potential selection issues. For instance, each of the three panels include only mothers who have (i) had more than one childbirth, (ii) moved from their birth state and (iii) not moved between pregnancies. 'Matched panel #2' and 'Matched panel #3' include only mothers who have not obtained additional education between pregnancies, and 'Matched panel #3' includes only mothers who have not changed their marital status (and, if married, their husband). The obvious concern is that these subsamples may

[11] These percentages are based on a sample of women born between 1957 and 1964, meaning that the time period of interest pre-dates the 1990–8 time period considered in this study.

be somehow unrepresentative of the overall population of mothers with multiple births due to the way in which the subsamples were selected.

To examine this issue, a random sample of mothers with multiple births was chosen from the natality data.[12] In particular, a random sample of 500 000 births was selected from the data. Table II reports the summary statistics for this random sample and each of the three matched panels. The variables are self-explanatory except perhaps for the 'Kessner index' variables. A value of 1 for the Kessner index indicates 'adequate' prenatal care, whereas a value of 3 indicates 'inadequate' prenatal care and 2 indicates 'intermediate' prenatal care. This index value is based upon information collected on prenatal care visits and gestation; a detailed description is provided in the appendix. For the non-indicator data items in Table II, standard deviations are reported in parentheses. Although the variable averages are similar between the random sample and the matched panels, there are some important differences in sample composition. The birth outcomes in the matched panels are better (higher birthweights, lower incidence of LBW, longer gestations) than those in the random sample. On average, the mothers in the panels are older, more educated, and more likely to be married and white. The women in the panels also smoke less and are more likely to obtain prenatal care earlier than those in the random sample.

Table II indicates a definite difference between the sample composition of the matched panels and the overall population. Our ultimate interest, however, lies in describing the conditional distribution of birth outcomes (such as birthweight) and analysing how smoking affects this conditional distribution. To get a sense of the conditional distributions of both birthweight and smoking participation (given other observables), Table III presents results from OLS regressions for the random sample and matched panels. The first four columns contain results using birthweight as the dependent variable, and the last four columns contain results using the smoking indicator as

Table II. Summary statistics

	Random sample	Matched panel #1	Matched panel #2	Matched panel #3
Birthweight (g)	3408 (566)	3431 (564)	3440 (553)	3457 (540)
LBW indicator	0.050	0.047	0.044	0.039
Gestation (wks)	39.07 (2.41)	39.17 (2.32)	39.21 (2.26)	39.25 (2.16)
Smoking indicator	0.179	0.158	0.155	0.130
# cigarettes/day (for smokers)	12.87 (8.05)	12.58 (7.97)	12.69 (7.95)	12.55 (7.83)
Age	28.20 (5.52)	28.61 (6.04)	28.44 (5.72)	28.48 (5.45)
Education (yrs)	12.74 (2.74)	13.55 (2.34)	13.65 (2.32)	13.88 (2.28)
Married indicator	0.758	0.803	0.826	0.869
Black indicator	0.179	0.138	0.106	0.074
Kessner index = 1	0.732	0.752	0.767	0.791
Kessner index = 2	0.206	0.193	0.184	0.169
Kessner index = 3	0.062	0.055	0.049	0.040
No prenatal visit	0.012	0.011	0.009	0.007
First prenatal visit in 1st trimester	0.804	0.823	0.837	0.859
First prenatal visit in 2nd trimester	0.153	0.138	0.129	0.113
First prenatal visit in 3rd trimester	0.031	0.027	0.025	0.020
Births	500 000	171 319	229 867	296 218
Mothers		82 595	110 260	141 929
Mothers with 2 births in sample		76 466	100 913	129 569

[12] As with the matched panels, only singleton births to white or black mothers are considered.

Table III. Comparison of OLS results from random and matched samples

	Dependent variable = birthweight				Dependent variable = smoking			
	Random sample	Panel #1	Panel #2	Panel #3	Random sample	Panel #1	Panel #2	Panel #3
Smoking	−248.19	−252.05	−248.11	−243.27				
	(2.14)	(3.99)	(3.45)	(3.20)				
Male	126.17	128.59	128.24	126.70				
	(1.51)	(2.57)	(2.18)	(1.88)				
Age	22.96	3.95	6.91	7.06	0.0245	0.0418	0.0359	0.0269
	(1.35)	(2.09)	(1.91)	(1.77)	(0.0009)	(0.0014)	(0.0013)	(0.0011)
Age^2	−0.405	−0.068	−0.122	−0.118	−0.00039	−0.00066	−0.00056	−0.00042
	(0.023)	(0.035)	(0.032)	(0.030)	(0.00001)	(0.00002)	(0.00002)	(0.00002)
High-school graduate	42.31	53.81	60.46	60.52	−0.0910	−0.1481	−0.1817	−0.1633
	(2.35)	(4.86)	(4.23)	(4.12)	(0.0018)	(0.0040)	(0.0035)	(0.0034)
Some college	74.41	90.78	99.90	91.34	−0.1653	−0.2488	−0.2865	−0.2614
	(2.72)	(5.32)	(4.55)	(4.52)	(0.0020)	(0.0041)	(0.0035)	(0.0035)
College graduate	102.97	105.35	111.06	100.89	−0.2508	−0.3305	−0.3652	−0.3286
	(2.99)	(5.69)	(4.91)	(4.73)	(0.0020)	(0.0041)	(0.0036)	(0.0036)
Married	59.98	53.42	50.10	64.43	−0.1471	−0.1660	−0.1611	−0.1912
	(2.27)	(4.20)	(3.66)	(3.65)	(0.0017)	(0.0032)	(0.0029)	(0.0029)
Black	−231.84	−243.54	−252.83	−252.04	−0.1279	−0.1155	−0.1176	−0.1252
	(2.52)	(4.43)	(4.24)	(4.36)	(0.0016)	(0.0027)	(0.0026)	(0.0027)
Kessner index = 2	−89.55	−108.01	−103.74	−100.93	0.0148	0.0207	0.0211	0.0166
	(3.24)	(5.66)	(4.80)	(4.19)	(0.0021)	(0.0034)	(0.0029)	(0.0024)
Kessner index = 3	−181.82	−223.21	−207.86	−176.48	0.0610	0.0671	0.0641	0.0511
	(6.39)	(12.22)	(11.02)	(10.20)	(0.0047)	(0.0087)	(0.0078)	(0.0073)
No prenatal visit	−85.45	−13.38	−17.48	−26.49	0.0628	0.0510	0.0707	0.0555
	(11.38)	(20.80)	(19.13)	(18.00)	(0.0077)	(0.0138)	(0.0127)	(0.0120)
First prenatal visit in 2nd trimester	74.43	92.60	91.16	89.12	0.0178	0.0215	0.0215	0.0210
	(3.64)	(6.42)	(5.54)	(4.96)	(0.0025)	(0.0042)	(0.0037)	(0.0032)
First prenatal visit in 3rd trimester	157.39	202.67	183.93	154.66	−0.0211	−0.0016	0.0043	0.0095
	(7.48)	(14.15)	(12.77)	(12.03)	(0.0058)	(0.0107)	(0.0097)	(0.0092)
R-squared	0.1071	0.1113	0.1076	0.0972	0.1078	0.1502	0.1662	0.1615
# observations	500 000	171 319	229 867	296 218	500 000	171 319	229 867	296 218

the dependent variable.[13] The model specification used in Table III, described in detail in Section 4, includes birth order effects, year effects and state effects (whose estimates are omitted from the table).[14]

Overall, the regression results are fairly similar for the random sample and the matched panels. Although some of the demographic variables vary in their impact on birthweight across the samples, the effect of smoking on birthweight remains remarkably stable. The estimates of the smoking indicator coefficient (first four columns of Table III) indicate an associated reduction in birthweight of 243–252 g.[15] In the birthweight regressions, the largest difference appears to be related to age. The gradient of the age effect is largest in the random sample, but the peak of the age profile for each of the four samples is extremely similar (between age 28 and

[13] A linear probability model for smoking participation gives nearly identical results to a probit model. This similarity arises since the fitted probabilities (of smoking) are in a rather narrow range.

[14] For the smoking regressions, the male indicator variable is omitted (though it has no effect on the results) since it is a birth outcome that should not be related to the smoking decision.

[15] The analogous regression results for LBW and gestation as dependent variables revealed the same degree of similarity.

age 30). For the smoking regressions, there are some differences in the education and prenatal care estimates.

None of the differences in Table III seem to suggest that selection on observables is causing a drastically different sample composition for the matched panels. On the other hand, it must still be acknowledged that selection on unobservables could be a factor.

3. THE INCONSISTENCY FROM INCORRECT MATCHING

This section considers the effect that incorrect matching has upon fixed effects estimation using matched panels. In particular, the potential inconsistency is compared with the original OLS (omitted variables) inconsistency that fixed effects estimation is intended to remove. To simplify the analysis and focus on the smoking effect γ, we consider the model in (1) without the additional covariates x_i. That is, for a given observed birth, we have

$$y_i = \gamma s_i + c_i + u_i \tag{2}$$

where the error disturbance consists of a 'mother fixed effect' c_i that is correlated with s_i and an idiosyncratic component u_i that is uncorrelated with s_i.

For this section, we also assume that each 'mother' has only two births, so that fixed effects estimation is equivalent to OLS estimation based upon differences of the matched pairs. For two births matched to each other (indexed i and j), the difference for the matched pair based upon (2) would be

$$y_i - y_j = \gamma(s_i - s_j) + (c_i - c_j) + (u_i - u_j) \tag{3}$$

Only those matched pairs with a change in smoking status ($s_i \neq s_j$) will have an effect on the estimate of γ. Correctly matched pairs have $c_i = c_j$, so that the pair difference eliminates the endogeneity from the mother fixed effect. (The assumption that the fixed effect is time-invariant will be revisited in Section 5.) For incorrectly matched pairs ('mismatches'), the fixed effects c_i and c_j may differ and impart an inconsistency into the estimation of γ.

Since smoking behaviour is positively correlated across pregnancies for a given mother, the likelihood of observing a change in smoking behaviour ($s_i \neq s_j$) depends upon whether the pair has been matched correctly. A mismatched pair will be more likely to have $s_i \neq s_j$, meaning that mismatched pairs will have a disproportionate effect on the fixed effects estimate of γ.

Let p_s denote the marginal probability of smoking in a given pregnancy. Then, the probability of a change in smoking behaviour for a mismatched pair (different mothers) is

$$\Pr(s_i \neq s_j | \text{mismatch}) = 2p_s(1 - p_s) \tag{4}$$

For a correctly matched pair, if i denotes the earlier birth and j denotes the later birth, the probability of a change in smoking behaviour is

$$\begin{aligned} \Pr(s_i \neq s_j | \text{match}) &= (1 - p_s)\Pr(s_j = 1 | s_i = 0) + p_s \Pr(s_j = 0 | s_i = 1) \\ &= (1 - p_s)(\delta_1 p_s) + p_s(\delta_2(1 - p_s)) \\ &= (\delta_1 + \delta_2)p_s(1 - p_s) \end{aligned} \tag{5}$$

where

$$\delta_1 \equiv \frac{\Pr(s_i = 0 | s_j = 1)}{\Pr(s_i = 0)} \quad \text{and} \quad \delta_2 \equiv \frac{\Pr(s_i = 1 | s_j = 0)}{\Pr(s_i = 1)}$$

Since smoking is positively correlated across pregnancies ($\delta_1 < 1,\ \delta_2 < 1$), the probability in (5) is lower than the probability in (4). For 'Matched panel #3', the joint probabilities of smoking in the two pregnancies are as follows: $\Pr(s_i = s_j = 0) = 0.821,\ \Pr(s_i = s_j = 1) = 0.081,\ \Pr(s_i = 0, s_j = 1) = 0.051$ and $\Pr(s_i = 1, s_j = 0) = 0.046$.[16] Although these probabilities are themselves affected by the presence of mismatches, they can still give a rough idea of relevant values for δ_1 and δ_2 in our application. If these joint probabilities reflected only correctly matched pairs, the implied values for the parameters in (5) would be $\delta_1 \approx 0.44$ and $\delta_2 \approx 0.42$.

Letting p_m denote the probability that a given pair is *mismatched*, an application of Bayes' law (using equations (4) and (5)) yields the conditional probability of a mismatch given a change in smoking behaviour:

$$\begin{aligned}\Pr(\text{mismatch} | s_i \neq s_j) &= \frac{2p_m p_s(1 - p_s)}{2p_m p_s(1 - p_s) + (\delta_1 + \delta_2)(1 - p_m)p_s(1 - p_s)} \\ &= \frac{2p_m}{2p_m + (\delta_1 + \delta_2)(1 - p_m)}\end{aligned}$$

The inconsistency of the fixed effects estimator is linked to this conditional probability:

$$\begin{aligned}\text{plim } \hat{\gamma}_{FE} &= \Pr(\text{mismatch} | s_i \neq s_j) E\left(\frac{y_i - y_j}{s_i - s_j} \middle| \text{mismatch}, s_i \neq s_j \right) \\ &\quad + \Pr(\text{match} | s_i \neq s_j) E\left(\frac{y_i - y_j}{s_i - s_j} \middle| \text{match}, s_i \neq s_j \right) \\ &= \gamma + \frac{2p_m}{2p_m + (\delta_1 + \delta_2)(1 - p_m)} [E(c|s = 1) - E(c|s = 0)] \qquad (6)\end{aligned}$$

Note that the term $E(c|s = 1) - E(c|s = 0)$ represents the usual omitted variables inconsistency (using OLS) for a regression having a single binary explanatory variable. Several interesting conclusions emerge from the inconsistency expression in (6).

1. Since smoking is positively correlated across pregnancies ($\delta_1 + \delta_2 < 2$), the mismatched pairs have a disproportionate effect on the inconsistency of the fixed effects estimator. That is,

$$\frac{2p_m}{2p_m + (\delta_1 + \delta_2)(1 - p_m)} > p_m$$

2. The inconsistency is increasing in p_m, the probability of a mismatched pair. For $p_m = 0$ (perfect matching), the inconsistency disappears. For $p_m = 1$ (all mismatches), the inconsistency is the same as the original omitted variables inconsistency. To visualize the effect of p_m on the inconsistency, Figure 2 plots the inconsistency factor (the fraction of the omitted variables inconsistency) as a function of p_m. Three curves are shown for $(\delta_1 + \delta_2)$ values in a reasonable

[16] These probabilities are based upon the first two births for any given 'matched' mother.

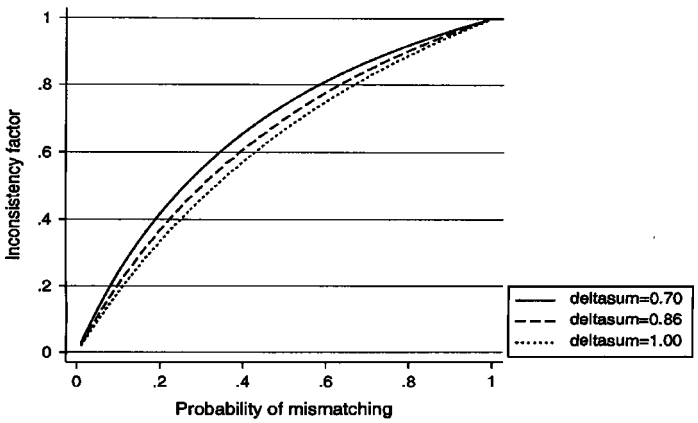

Figure 2. Effect of mismatching on fixed effects inconsistency

range for the smoking application, with the middle curve ($\delta_1 + \delta_2 = 0.86$) corresponding to the implied values discussed above.

3. The inconsistency is decreasing in $(\delta_1 + \delta_2)$. When $\delta_1 + \delta_2 = 2$ (as would be the case if smoking were uncorrelated across pregnancies), the mismatched pairs have a proportional effect on the inconsistency. Though not reasonable in the present setting, notice that $\delta_1 + \delta_2 > 2$ (negative correlation in smoking behaviour) would imply that mismatched pairs would have *less* than a proportional effect on the inconsistency.

The curves in Figure 2 stress the importance of correctly matching births together. Looking at the middle curve, a mismatch rate of only 10% results in an inconsistency equal to 20% of the omitted variables inconsistency. Matters become even worse for higher mismatch rates: a mismatch rate of 20% has an inconsistency factor of 36.8%, and a mismatch rate of 30% has an inconsistency factor of 49.9%.

The basic message from this section is that incorrect matching will tend to bias the fixed effects estimates of γ in the direction of the OLS estimator. The fixed effects estimates of γ reported in the next section turn out to be smaller in magnitude than the OLS estimator. To the extent that incorrect matching is causing a bias towards the OLS estimator, the smaller estimated magnitudes should actually be interpreted as still being too large in magnitude. In the next section, a variable that proxies for correct matching will be used to assess the inconsistency that arises from incorrect matching in the fixed effects framework.

Although not pursued here, it would be interesting to extend the results of this section to other situations, such as a continuous explanatory variable or allowing the mismatch rate to explicitly depend on observables.

4. EMPIRICAL RESULTS

4.1. Fixed Effects Specifications

The fixed effects version of the model in (1) is written as

$$y_{ib} = x_{ib}'\beta + \gamma s_{ib} + c_i + u_{ib} \quad (i = 1, \ldots, n; b = 1, \ldots, B_i) \tag{7}$$

Mothers (or, more accurately, matched units) are indexed by i, and births are indexed by b (with the total number of births per mother $B_i \geq 2$). The dependent variable y denotes the birth outcome of interest, which will be either (i) birthweight (in grams), (ii) an LBW indicator variable (equal to one if birthweight is less than 2500 g), or (iii) gestation (in weeks). The x_{ib} vector contains birth-specific explanatory variables (other than smoking), and s_{ib} is an indicator for smoking during pregnancy. The fixed effect c_i is allowed to be correlated with the observables x_{ib} and s_{ib} but is restricted to be time-invariant. It is conceivable that the unobservable mother fixed effect varies over time and possibly in a way that is correlated with changes in smoking behaviour. This possibility is considered in more detail in Section 5.

Linear regression models are used for each of the three dependent variables. For the LBW indicator variable, the model estimated is a linear probability model; this model does not have its usual drawbacks in the current setting since the conditional probabilities of LBW are always quite low (so that the linear approximation is suitable). The model specification includes variables associated with mother's age, education, marital status, race and prenatal care. The education variables are indicators for high-school graduates (years of education = 12), women who attended some college (years of education = 13, 14 or 15) and college graduates (years of education $\geq$ 16). In order to interpret estimates on the categorical variables, note that the omitted categories are less than 12 years of education, Kessner index = 1 ('adequate' prenatal care) and first prenatal visit occurring during the first trimester. The model specification also includes birth order effects (indicator variables for each possible parity), year effects and state effects.

Table IV reports the results from OLS and fixed effects regressions for each of the three matched panels when the dependent variable is birthweight (in grams).[17] The first six columns consider the model specification described above with only a smoking indicator variable. The last six columns consider a specification where number of cigarettes (smoked per day) is also included as an explanatory variable. The analogous results for the LBW and gestation regressions (Table XIII and XIV, respectively) are provided in the appendix.

Consistent with the basic omitted variables story, the fixed effects estimates of smoking on birthweight are much lower than the OLS estimates. For 'Matched panel #3', there is nearly a 100 g difference between the OLS estimate (−243.38 g) and the fixed effects estimate (−144.07 g); the difference is 65–70 g for the other two panels. Note that the larger difference for 'Matched panel #3' is also consistent with the theory on incorrect matching developed in Section 3. If 'Matched panels #1 and #2' have more incorrect matches, we would expect to see their associated fixed effects estimates being closer to the OLS estimates. When the number of cigarettes is included as an explanatory variable in the birthweight regression, both the level of the smoking effect and the gradient of the effect (with respect to number of cigarettes) are found to have lower magnitudes in the fixed effects regressions.

For the LBW indicator and gestation regressions (reported in the appendix), the results are qualitatively similar to those obtained in Table IV. For both dependent variables, the magnitude of the smoking effect is estimated to be smaller in the fixed effects specification for each of the three matched panels. The fixed effects estimates of the effect of smoking on incidence of LBW are 36–43% lower than the OLS estimates. For instance, the fixed effects estimate for 'Matched panel #3' estimates an increase in LBW incidence of 2.43% associated with smoking, as compared to the increase of 4.29% implied by OLS. The fixed effects estimates also imply a smaller effect

[17] A random effects specification (c uncorrelated with x and s) was also considered, but not surprisingly was rejected overwhelmingly by statistical tests.

Table IV. Matched panel regressions (dependent variable = birthweight)

	Panel #1		Panel #2		Panel #3		Panel #1		Panel #2		Panel #3	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
Smoking	−252.05	−178.19	−248.11	−177.62	−243.27	−144.04	−193.49	−136.86	−189.84	−129.37	−176.35	−98.34
	(3.99)	(5.81)	(3.45)	(4.97)	(3.20)	(4.75)	(6.91)	(8.88)	(5.94)	(7.61)	(5.62)	(7.04)
# cigarettes							−5.04	−3.77	−4.93	−4.34	−5.70	−4.28
							(0.45)	(0.57)	(0.38)	(0.48)	(0.37)	(0.45)
Male	128.59	126.81	128.24	132.16	126.70	133.58	129.12	127.02	128.36	131.80	126.87	133.49
	(2.57)	(3.02)	(2.18)	(2.56)	(1.88)	(2.08)	(2.57)	(3.04)	(2.18)	(2.57)	(1.89)	(2.09)
Age	3.95	−15.31	6.91	−15.51	7.06	−15.98	5.27	−15.67	8.11	−15.34	8.22	−15.49
	(2.09)	(5.29)	(1.91)	(4.59)	(1.77)	(3.96)	(2.10)	(5.32)	(1.92)	(4.62)	(1.78)	(3.98)
Age^2	−0.068	0.258	−0.122	0.212	−0.118	0.317	−0.087	0.258	−0.139	0.211	−0.136	0.308
	(0.035)	(0.069)	(0.032)	(0.061)	(0.030)	(0.053)	(0.035)	(0.070)	(0.032)	(0.061)	(0.030)	(0.053)
High-school graduate	53.81	23.70	60.46		60.52		50.13	23.11	57.05		57.39	
	(4.86)	(9.54)	(4.23)		(4.12)		(4.89)	(9.64)	(4.26)		(4.15)	
Some college	90.78	55.74	99.90		91.34		86.42	56.56	95.39		87.15	
	(5.32)	(11.66)	(4.55)		(4.52)		(5.35)	(11.78)	(4.58)		(4.56)	
College graduate	105.35	73.97	111.06		100.89		100.16	75.38	106.18		96.30	
	(5.69)	(14.80)	(4.91)		(4.73)		(5.72)	(14.91)	(4.94)		(4.77)	
Married	53.42	48.40	50.10	40.64	64.43		52.76	46.55	49.12	38.60	63.16	
	(4.20)	(5.88)	(3.66)	(5.08)	(3.65)		(4.22)	(5.93)	(3.68)	(5.12)	(3.67)	
Black	−243.54		−252.83		−252.04		−246.29		−255.73		−255.71	
	(4.43)		(4.24)		(4.36)		(4.44)		(4.25)		(4.38)	
Kessner index = 2	−108.01	−102.24	−103.74	−92.56	−100.93	−84.43	−107.91	−102.10	−103.32	−93.03	−100.53	−84.24
	(5.66)	(6.30)	(4.80)	(5.35)	(4.19)	(4.45)	(5.68)	(6.34)	(4.82)	(5.38)	(4.20)	(4.48)
Kessner index = 3	−223.21	−200.75	−207.86	−175.87	−176.48	−143.91	−222.96	−199.86	−205.46	−174.26	−174.78	−143.96
	(12.22)	(13.10)	(11.02)	(11.50)	(10.20)	(10.28)	(12.31)	(13.26)	(11.10)	(11.63)	(10.28)	(10.39)
No prenatal visit	−13.38	−36.70	−17.48	−34.18	−26.49	−42.35	−9.49	−41.53	−15.18	−37.78	−22.27	−39.64
	(20.80)	(20.49)	(19.13)	(18.22)	(18.00)	(16.57)	(20.99)	(20.77)	(19.30)	(18.45)	(18.16)	(16.76)
First prenatal visit in 2nd trimester	92.60	81.17	91.16	75.77	89.12	66.56	92.85	79.79	91.54	75.98	89.31	66.24
	(6.42)	(7.21)	(5.54)	(6.19)	(4.96)	(5.27)	(6.45)	(7.26)	(5.56)	(6.24)	(4.98)	(5.30)
First prenatal visit in 3rd trimester	202.67	180.33	183.93	142.88	154.66	111.90	202.74	177.70	182.81	140.21	153.97	112.16
	(14.15)	(15.77)	(12.77)	(13.88)	(12.03)	(12.49)	(14.26)	(15.96)	(12.86)	(14.04)	(12.13)	(12.63)

of smoking upon length of gestation. The statistical difference in the smoking indicator gestation regressions is smaller than found in the birthweight regressions. In addition, when number of cigarettes is included in the gestation regressions, many of the fixed effects smoking estimates become statistically insignificant.

As discussed previously, the presence of incorrect matching suggests that the fixed effects results described above may still be prone to a significant amount of omitted variables inconsistency. The next subsection uses a proxy for correct matching in order to assess the extent of this inconsistency.

4.2. Using a Proxy for Correct Matching

To gauge the degree of incorrect matching and its effect on the estimates reported in the previous subsection, we utilize a proxy for correct matches that is available for the earlier part of the sampling period. In particular, for most births occurring between 1990 and 1994, the natality data has a variable with the 'interval since last live birth'. For any birth beyond a mother's first child, the possible values for this variable are as follows: 1–11 months, 12–17 months, 18–23 months, 24–35 months, 36–47 months, 48–59 months, 60–71 months and 72 months or longer. For a given matched pair of births, the variable *proxy* is defined as follows:

$$proxy = \begin{cases} 1 & \text{if observed interval agrees with interval since last birth record} \\ 0 & \text{if observed interval disagrees with interval since last birth record} \end{cases} \quad (8)$$

This indicator variable is only defined when both births in a matched pair occur between 1990 and 1994. A value of zero for *proxy* is extremely strong evidence of an incorrect match since only miscoding of the interval record or birth month could result in $proxy = 0$ for a correct match.

Table V reports descriptive statistics of *proxy* for each of the three matched panels. 'Matched panels #1 and #2' both have a large degree of incorrect matching, with approximately one-third of the observations having $proxy = 0$. 'Matched panel #3', which requires a match on marital status and father's age, has $proxy = 0$ for only 18.8% of the subsample. These percentages (last column of Table V) should be viewed as lower bounds on the mismatch rate since pairs with $proxy = 1$ could also represent incorrect matches.

To estimate the amount of additional mismatching, we consider the frequencies of births that occur just outside of the recorded interval since last birth. For instance, in 'Matched panel #3' with a recorded 12–17 month interval, the percentage of births at 11 months and 18 months is 0.54% and 0.55%, respectively. To be conservative, the larger percentage (0.55%) is used and multiplied by the size of the interval (six months) to estimate that an additional 3.30% of the 12–17 month interval observations is mismatched. Using this approach and aggregating over the possible recorded intervals, the amount of additional mismatching is estimated to be 7.3% for 'Matched panel #1', 7.9% for 'Matched panel #2' and 5.7% for 'Matched panel #3'. Taken

Table V. Descriptive statistics for *proxy*

	Number of observations with recorded interval	Percentage with $proxy = 1$	Percentage with $proxy = 0$
Matched panel #1	27 908	66.4%	33.6%
Matched panel #2	38 884	65.5%	34.5%
Matched panel #3	49 490	81.2%	18.8%

together with the *proxy* values, these estimates suggest an overall mismatch rate of 40–43% for 'Matched panels #1 and #2' and an overall mismatch rate of 24–25% for 'Matched panel #3'.

In light of the theoretical analysis in Section 3, the high mismatch rates suggest that the fixed effects results (such as those in Table IV) are still affected by a large amount of omitted variables inconsistency. Before using the proxy variable to assess this consistency, we first consider some rough calculations based upon the mismatch rates and the inconsistency factors from Figure 2. The high mismatch rates for 'Matched panels #1 and #2' correspond to somewhere near a 60% inconsistency factor according to Figure 2. Taking the OLS estimate of the smoking effect to be −250 g and the fixed effects estimates for these two panels to be roughly −180 g (see Table IV), the implied 'true' fixed effects estimate (with the inconsistency removed) would be around −75 g. The same calculation applied to 'Matched panel #3' (using an inconsistency factor of about 40% from Figure 2) also yields a 'true' fixed effects estimate of around −75 g.

Now turning to the proxy variable, Table VI provides a summary of the smoking effect results (for birthweight, LBW and gestation) for each of the matched panels. The model specification is the same as described in Section 4.1. For each panel, several different estimation samples are considered. First, the full sample results (of Section 4.1) are provided. Then, in the earlier birth pairs for which the *proxy* variable is available, results are provided separately for the $proxy = 1$ subsample and the $proxy = 0$ subsample. Since *proxy* is not available for later observations, we

Table VI. Summary of the effects of smoking on birth outcomes

	Random sample	Matched panel #1		Matched panel #2		Matched panel #3	
	OLS	OLS	FE	OLS	FE	OLS	FE
Effect on birthweight							
Full sample	−248.19	−252.05	−178.19	−248.11	−177.62	−243.27	−144.04
	(2.14)	(3.99)	(5.81)	(3.45)	(4.97)	(3.20)	(4.75)
Early subsample		−260.62	−73.83	−259.45	−78.88	−247.01	−66.99
($proxy = 1$)		(8.67)	(14.50)	(7.00)	(11.74)	(5.87)	(9.97)
Early subsample		−244.27	−220.63	−244.65	−239.54	−234.20	−213.62
($proxy = 0$)		(11.88)	(16.53)	(9.41)	(12.97)	(11.33)	(15.79)
Later subsample		−248.72	−117.93	−318.64	−166.91	−245.03	−120.90
($\widehat{proxy} > 0.80$)		(12.42)	(18.93)	(15.56)	(25.69)	(6.54)	(10.06)
Effect on LBW							
Full sample	0.0492	0.0485	0.0308	0.0451	0.0294	0.0429	0.0243
	(0.0011)	(0.0019)	(0.0025)	(0.0017)	(0.0021)	(0.0015)	(0.0020)
Early subsample		0.0480	0.0073	0.0485	0.0134	0.0412	0.0120
($proxy = 1$)		(0.0042)	(0.0067)	(0.0035)	(0.0055)	(0.0028)	(0.0045)
Early subsample		0.0413	0.0289	0.0468	0.0399	0.0469	0.0452
($proxy = 0$)		(0.0057)	(0.0065)	(0.0046)	(0.0052)	(0.0055)	(0.0064)
Later subsample		0.0508	0.0232	0.0626	0.0403	0.0424	0.0201
($\widehat{proxy} > 0.80$)		(0.0063)	(0.0093)	(0.0081)	(0.0202)	(0.0031)	(0.0042)
Effect on gestation							
Full sample	−0.152	−0.188	−0.096	−0.154	−0.056	−0.136	−0.080
	(0.010)	(0.019)	(0.028)	(0.016)	(0.023)	(0.015)	(0.023)
Early subsample		−0.205	0.065	−0.161	0.148	−0.110	0.047
($proxy = 1$)		(0.042)	(0.075)	(0.034)	(0.061)	(0.027)	(0.051)
Early subsample		−0.077	−0.055	−0.133	−0.111	−0.117	−0.126
($proxy = 0$)		(0.054)	(0.073)	(0.042)	(0.057)	(0.052)	(0.071)
Later subsample		−0.211	−0.130	−0.309	−0.401	−0.156	−0.177
($\widehat{proxy} > 0.80$)		(0.062)	(0.097)	(0.078)	(0.128)	(0.029)	(0.046)

attempted to deal with some of the incorrect matching inconsistency by forming a predicted *proxy* (denoted $\widehat{proxy}$) for the later subsample. In particular, a first-stage probit regression was estimated on the *proxy* subsample (using *proxy* as the dependent variable and the maternal characteristics used for matching as independent variables).[18] Out-of-sample fitted values $\widehat{proxy}$ for the later birth pairs were based upon the probit regression, and Table VI reports results for observations in the later subsample for which $\widehat{proxy} > 0.80$.[19]

Our discussion will focus on the birthweight results contained in the top panel of Table VI. The most striking results are those for the $proxy = 1$ and $proxy = 0$ earlier subsamples. The $proxy = 1$ sample is the most reliable of the samples considered since the birth pairs are likely to be correctly matched, whereas the $proxy = 0$ birth pairs are almost certainly mismatched. As expected, for the incorrectly matched $proxy = 0$ sample, the fixed effects estimates of the smoking effect are quite similar to the OLS estimates. On the other hand, the $proxy = 1$ fixed effects estimates (between -79 g and -67 g) are much smaller than the overall fixed effects estimates based upon the full samples. Interestingly, these estimates are quite close to the rough calculation of -75 g described above. For the later subsample, the use of a fitted proxy seems to remove some of the incorrect matching inconsistency of the overall sample for the first and third panels. These results are somewhat less compelling than the $proxy = 1$ subsample since even this subsample is likely to still contain a significant fraction of incorrect matches. The estimated effects on LBW are qualitatively similar to those for birthweight. For gestation, however, the $proxy = 1$ results cast doubt on the fixed effects results from the overall sample that had indicated a significant negative effect of smoking.

4.3. Heterogeneity of Smoking Effects

The previous studies of smoking's effect on birth outcomes (cited in the Introduction) have treated the smoking effect as homogeneous. That is, the parameter γ in the model (1) is assumed not to vary over i. The fixed effects model introduced in (7) maintains the same homogeneous assumption on the smoking effect. However, if the smoking effect is heterogeneous, the estimate of γ under a homogeneity assumption (whether in an instrumental variables framework or a panel data framework) is not necessarily a consistent estimate of the average smoking effect.

The fixed effects approach results in relatively low standard errors on the effects of smoking, especially compared to instrumental variables studies. As a practical matter, then, there is hope of identifying heterogeneity in the smoking effects using the fixed effects approach. One way to incorporate heterogeneity is to explicitly allow the smoking effect to depend upon observables, rewriting the fixed effects model (7) as

$$y_{ib} = x'_{ib}\beta + \gamma_{ib}s_{ib} + c_i + u_{ib} \tag{9}$$

The heterogeneous smoking effect γ_{ib} (which may vary over mothers *and* births) is assumed to be a linear function of observables z_{ib} (including a constant) plus a random disturbance:

$$\gamma_{ib} = z'_{ib}\rho + e_{ib},\ E(e_{ib}|x_i, z_i, s_i) = 0 \tag{10}$$

[18] The probit specification included birth order dummies, education level dummies (rather than categories), age, age squared, race and marital status. It is important to use dummies for education since less frequent education levels (like 13 or 15 years of education) are easier to match correctly.

[19] Other cutoff levels give similar results, but choosing the cutoff level much higher than 0.80 results in a much smaller sample with imprecise estimates.

Combining (9) and (10) yields

$$y_{ib} = x'_{ib}\beta + (z_{ib}s_{ib})'\rho + c_i + (s_{ib}e_{ib} + u_{ib}) \tag{11}$$

The heterogeneity of the smoking effect is picked up by the interaction terms in (11). Note that components of ρ can even be identified for birth-invariant z_{ib} variables (such as mother's race) since the interaction with s_{ib} allows for variation of the interaction variable. Both β and ρ can be estimated with the fixed effects estimator. If a particular component of ρ is found to be statistically significant, this finding would be consistent with the smoking effect varying over different values of the associated covariate.

The results from fixed effects estimation of the model in (11) are reported in Table VII. The model specification is the same as considered in Table IV (with birth order, year and state effects), except that the smoking indicator is also interacted with each of the other explanatory variables.[20] Only the results for 'Matched panel #3' are reported, with estimates provided for the full sample and the $proxy = 1$ subsample. Also, to focus on the issue of heterogeneity, only the estimates for the interaction variables are reported in Table VII. The coefficient estimates provide some evidence of heterogeneity. The quadratic age specification suggests that the smoking effect is

Table VII. Interaction fixed effects regression results (panel #3)

	Full sample	Early subsample ($proxy = 1$)
Interaction of smoking indicator with:		
Constant	188.52	317.75
	(89.12)	(182.16)
Male	−0.80	2.88
	(6.20)	(10.69)
Age	−21.86	−30.32
	(6.80)	(13.97)
Age^2	0.33	0.56
	(0.12)	(0.25)
High-school graduate	−31.94	−6.65
	(13.32)	(26.89)
Some college	−14.81	30.44
	(16.53)	(34.30)
College graduate	70.27	144.65
	(21.82)	(41.92)
Married	25.58	−22.80
	(11.30)	(26.12)
Black	79.70	73.15
	(17.76)	(34.78)
Kessner index = 2	−7.07	−5.21
	(12.21)	(21.12)
Kessner index = 3	−40.43	29.09
	(23.10)	(39.32)
No prenatal visit	26.99	25.75
	(34.80)	(58.42)
First prenatal visit in 2nd trimester	−9.50	−9.81
	(13.42)	(23.02)
First prenatal visit in 3rd trimester	3.13	19.31
	(27.55)	(46.09)

[20] The inclusion of additional interactions with birth order, year and/or state effects has basically no effect on the results.

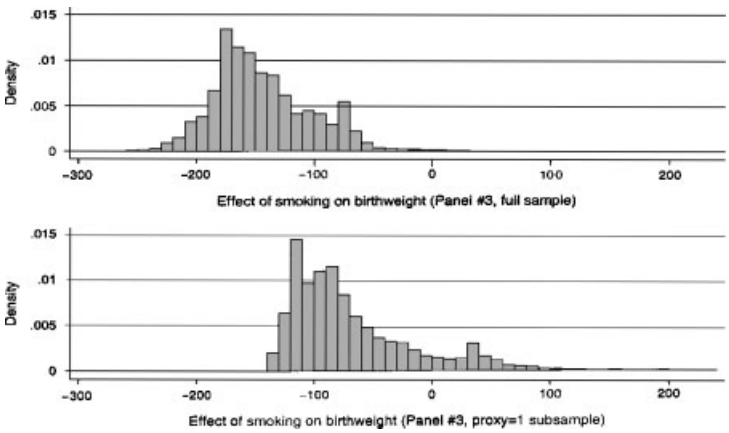

Figure 3. Histograms of (in-sample) fitted smoking effects

most pronounced (most negative) around age 30. The effect of smoking is lower for college graduates than for less than high school educated mothers (a difference of 70 g in the full sample and 145 g in the $proxy = 1$ subsample). The effect is also lower for blacks than whites (70–80 g difference). Nearly all of the other interaction variables are estimated to be statistically insignificant (at a 5% level).

Of course, it is possible that some of the effects found in Table VII may result from some other misspecification of the model (e.g., omitted nonlinearities or interactions of other covariates). Given this caveat and the rather high standard errors of the individual estimates (due to the inherent multicollinearity of the interaction variables), an alternative approach for gauging the extent of heterogeneity is to view the 'fitted smoking effects' for the sample observations. The fitted smoking effects are given by $z'_{ib}\hat{\rho}$ and can be computed directly from the estimates in Table VII. Figure 3 provides histograms (one for the full sample and one for the $proxy = 1$ subsample) of the fitted smoking effects. Recall that the homogeneous smoking effects previously estimated were -144 g for the full sample and -67 g for the $proxy = 1$ subsample. While these estimates seem to capture the overall level of the effects pretty well, the histograms suggest that there is considerable dispersion in the effects as well. The standard deviations associated with the two histograms are 50.9 g and 40.5 g, respectively. Interestingly, the histogram for the $proxy = 1$ subsample appears to be right-skewed and has no fitted effects below -140 g. For some portion of this sample, the effect of smoking appears to be negligible. (Although some of the fitted effects are actually *positive*, these values are not significantly different from zero at any reasonable statistical level.)

5. POSSIBLE VIOLATIONS OF STRICT EXOGENEITY

It is well known (e.g., Wooldridge (2002, chap 10.5)) that the following *strict exogeneity* assumption ensures consistency of the fixed effects estimator for the model in (7):

$$E(u_{ib}|x_{i1}, \ldots, x_{i_{B_i}}, s_{i1}, \ldots, s_{i_{B_i}}, c_i) = 0 \tag{12}$$

This assumption is often made in empirical work without much critical analysis. Since violations of this assumption can cause inconsistency of the estimated causal effect, it is important to think

about the direction of the inconsistency caused by possible violations. This section discusses three possible reasons that strict exogeneity would not hold: (1) a feedback effect (whereby smoking behaviour can be influenced by prior birth outcomes), (2) a correlation between changes in smoking status and changes in unobserved variables, and (3) misclassification of smoking status. To simplify exposition, we assume for the remainder of this section that each mother has only two births (that is, $B_i = 2$ for each i) and that the y variable is birthweight. Omitting the subscript i, the strict exogeneity assumption can then be rewritten:

$$E(u_1|x_1, x_2, s_1, s_2, c) = E(u_2|x_1, x_2, s_1, s_2, c) = 0 \tag{13}$$

5.1. Feedback Effect

It is possible that the outcome of a mother's first pregnancy may affect behaviour during her second pregnancy. For instance, a particularly low birthweight for the first birth (associated with a 'bad draw' of u_1) might cause a mother to quit smoking for her second pregnancy. Likewise, a particularly high birthweight for the first birth ('good draw' of u_1) might cause a smoking mother to continue smoking or even cause a non-smoking mother to adopt smoking. (Note that a 'non-smoking mother' could be a woman who had smoked in the past but quit smoking for her first pregnancy.) In terms of the model notation above, this feedback effect story would imply that $E[s_2u_1|s_1] > 0$, in violation of (13). The implication is that $E[(s_2 - s_1)(u_2 - u_1)] < 0$, meaning that this violation would likely lead to a negative bias in the fixed effects estimate of γ (the coefficient on s).

One way to deal with this violation of strict exogeneity would be to consider the sample of mothers for whom at least *three* births are available. For mothers with three births, smoking during the first pregnancy could be used as an instrument for the change in smoking behaviour during the second and third pregnancies. This instrument is likely to be valid in the sense that it should not be correlated with the birth-specific disturbances in the second and third pregnancies. Unfortunately, the matched panels do not contain enough observations to make this estimation strategy feasible. Considering 'Matched panel #3', the number of mothers with at least three births is 12 360. Although there is a statistically significant correlation between the proposed instrument (smoking in a prior pregnancy) and the change in smoking during subsequent pregnancies, the resulting instrumental variables estimates are extremely imprecise due to the relatively small sample size. In an IV regression of birthweight difference on only smoking status difference (omitting other covariates), for instance, the standard error on the smoking effect is nearly 370 g. This impression suggests that the IV approach would require a much larger sample size to empirically identify smoking's effect on birthweight.

5.2. Correlated Changes in Maternal Behaviour

If a woman smokes during her first pregnancy and decides to quit for her second pregnancy (either due to the 'feedback effect' described above or for some other reason), it may be the case that she alters other unobservable behaviours in a way that positively influences birthweight in the second pregnancy. Likewise, women who adopt smoking may also adopt other unobservable negative impact behaviours. There are several ways of thinking about this story in the context of the fixed effects model. One way is to view the mother's fixed effect c as being birth-varying (time-varying). Another way is to view c as being unchanged from the first to the second birth and to include

Table VIII. Changes in prenatal care (panel #3, $proxy = 1$ subsample)

	Mothers who quit smoking (%)		Mothers who adopt smoking (%)	
	1st birth	2nd birth	1st birth	2nd birth
First prenatal visit occurs:				
in 1st trimester	78.1	80.7	76.2	73.8
in 2nd trimester	16.9	15.0	19.0	19.7
in 3rd trimester	3.6	3.2	3.6	4.9
never	1.4	1.1	1.2	1.7
Kessner index value:				
1 ('adequate care')	71.1	73.5	68.9	65.4
2 ('intermediate care')	22.3	20.4	24.5	25.5
3 ('inadequate care')	6.6	6.1	6.6	9.1

any change in the unobservable maternal behaviour in u_2. Adopting this latter view, the argument above would imply that $E[(s_2 - s_1)u_2] < 0$, in violation of (13). As with the feedback effect, the implication is that $E[(s_2 - s_1)(u_2 - u_1)] < 0$, which again would lead to a negative bias in the fixed effects estimate of γ.

The observable variables in the natality data allow one to partially deal with correlated changes in maternal behaviour. These observables also provide evidence that mothers who change their smoking behaviour also change other prenatal care in the 'same direction'. Considering the $proxy = 1$ subsample of 'Matched panel #3' (the most reliable in terms of correct matching), Table VIII reports descriptive statistics on the prenatal care variables for both women who quit smoking and women who adopt smoking for the second pregnancy. Mothers who quit smoking are more likely to have a prenatal visit during their first trimester and more likely to be coded as having 'adequate' prenatal care. In contrast, mothers who adopt smoking are less likely to have a first trimester prenatal visit and 'adequate' care. These results suggest that other unobservables (such as drinking or poor nutrition) would behave similarly, leading to the negative bias described above. Of course, without more detailed data on maternal prenatal behaviour, it is not possible to quantify the magnitude of this bias.

We should point out that another story would imply a bias going in the opposite direction. In particular, mothers who quit smoking may feel that they can adopt other unobservable negative impact activities in place of smoking. This type of behaviour would be analogous to the 'compensating behaviour' theory considered by Peltzman (1975) in the context of seatbelt usage and other driving behaviours. The results of Table VIII are inconsistent with this theory, and the theory also seems less convincing for unobservable behaviour.

5.3. Misclassification of Smoking Status

It is well known that the fixed effects estimator can exacerbate the bias caused by the measurement error of explanatory variables (e.g., Griliches and Hausman, 1986). Freeman (1984) and Jakubson (1986) consider the specific case of measurement error (or 'misclassification') of a binary explanatory variable. Both studies consider misclassification of a binary union status variable in the context of a wage regression and find that misclassification causes a more severe attenuation bias for the fixed effects estimator than the cross-sectional estimator. The basic idea is that misclassification causes the measurement error of changes in union status to be more severe than the measurement error of union status itself.

In this section, we consider the effect of misclassification of smoking status on the fixed effects estimator. There are two main differences from the framework considered by Freeman (1984) and Jakubson (1986). First, we explicitly consider the case where the misclassification is one-sided. In particular, we allow for the possibility that smokers misclassify themselves as non-smokers but not the possibility that non-smokers misclassify themselves as smokers. Second, we allow for misclassification in the second time period to depend upon misclassification in the first time period. That is, we allow for the possibility that a woman who smoked during her first pregnancy and misclassified herself as a non-smoker is more likely to do the same in her second pregnancy.

To focus attention on the smoking status variable, we consider a simplified version of the panel data model in which the other explanatory variables x are omitted:

$$y_1 = \gamma s_1^* + c + u_1 \tag{14}$$

$$y_2 = \gamma s_2^* + c + u_2 \tag{15}$$

where s_1^* and s_2^* denote the true smoking status in each of the two pregnancies. Let s_1 and s_2 denote the reported smoking status in each of the two pregnancies. As discussed above, assume that non-smokers do not misclassify themselves as smokers: $\Pr(s_1 = 1|s_1^* = 0) = \Pr(s_2 = 1|s_2^* = 0) = 0$. The misclassification probabilities for smokers are defined as follows:

$$p = \Pr(s_1 = 0|s_1^* = 1) = \Pr(s_2 = 0|s_1^* = 0, s_2^* = 1) \tag{16}$$

$$q = \Pr(s_2 = 0|s_1^* = 1, s_1 = 0, s_2^* = 1) \tag{17}$$

$$r = \Pr(s_2 = 0|s_1^* = 1, s_1 = 1, s_2^* = 1) \tag{18}$$

The misclassification probability p is the probability that a smoker misclassifies herself in her first smoking pregnancy. The misclassification probability q is the probability that a smoker misclassifies herself in a second smoking pregnancy conditional on misclassifying herself in her first smoking pregnancy. The probability r is the second pregnancy misclassification probability conditional on correct classification in the first smoking pregnancy. One would suspect that these three probabilities satisfy the inequalities $r < p < q$. Based upon the definitions in (16)–(18), Table IX gives the conditional probabilities of the four possible reported status pairs given the value of the true status pair.

In order to determine the bias associated with misclassification, the unconditional probabilities of the true status pairs must also be specified:

$$v_{0,0} \equiv \Pr(s_1^* = 0,\ s_2^* = 0)$$

$$v_{0,1} \equiv \Pr(s_1^* = 0,\ s_2^* = 1)$$

Table IX. Conditional probabilities of reported status given true status ($\Pr((s_1, s_2)|(s_1^*, s_2^*))$)

	$s_1 = 0,\ s_2 = 0$	$s_1 = 0,\ s_2 = 1$	$s_1 = 1,\ s_2 = 0$	$s_1 = 1,\ s_2 = 1$
$s_1^* = 0,\ s_2^* = 0$	1	0	0	0
$s_1^* = 0,\ s_2^* = 1$	p	$1 - p$	0	0
$s_1^* = 1,\ s_2^* = 0$	p	0	$1 - p$	0
$s_1^* = 1,\ s_2^* = 1$	pq	$p(1 - q)$	$(1 - p)r$	$(1 - p)(1 - r)$

$$v_{1,0} \equiv \Pr(s_1^* = 1,\ s_2^* = 0)$$
$$v_{1,1} \equiv \Pr(s_1^* = 1,\ s_2^* = 1)$$

For the cross-sectional case, the probability limit of the OLS estimator (assuming misclassification is the only source of endogeneity) is

$$\text{plim}\ \hat{\gamma}_{OLS} = \frac{\gamma}{1+p} \tag{19}$$

For the panel data case, the probability limit of the fixed effects estimator (again assuming misclassification is the only source of endogeneity) is

$$\text{plim}\ \hat{\gamma}_{FE} = \gamma \left[\frac{f}{1 + r\dfrac{v_{1,1}}{v_{1,0}}} + \frac{1-f}{1 + \dfrac{p(1-q)v_{1,1}}{(1-p)v_{0,1}}} \right] \tag{20}$$

where

$$f = \Pr(s_1 = 1,\ s_2 = 0 | s_1 \neq s_2) = \frac{(1-p)(v_{1,0} + rv_{1,1})}{(1-p)(v_{1,0} + v_{0,1} + rv_{1,1}) + p(1-q)v_{1,1}} \tag{21}$$

The first term within the brackets in (20) corresponds to the attenuation bias caused by mothers who report cessation of smoking ($s_1 = 1,\ s_2 = 0$), and the second term corresponds to the attenuation bias caused by mothers who report adoption of smoking ($s_1 = 0,\ s_2 = 1$). The attenuation bias depends upon the misclassification probabilities (p, q and r), the probability f, and the ratios $v_{1,1}/v_{0,1}$ and $v_{1,1}/v_{1,0}$. While there are no direct estimates available for any of these quantities, there is some suggestive evidence for some of them. Table X reports frequencies of reported status pairs from two states (Missouri and Washington) for which this information was available.[21] The numbers suggest a value of f in the 0.4–0.5 range and values of the ratios $v_{1,1}/v_{0,1}$ and $v_{1,1}/v_{1,0}$ in the 2.3–4.3 range.

Table X. State level information on reported smoking status

Smoked during first pregnancy	Smoked during second pregnancy	Missouri (1989–1997)	Washington (1990–1998)
No	No	77.5%	80.5%
No	Yes	4.9%	4.4%
Yes	No	3.3%	4.5%
Yes	Yes	14.3%	10.6%

Sources: For Missouri, the percentages come from *Missouri Monthly Vital Statistics*, Vol. 35, No. 9. For Washington, the percentages come from data obtained by the author from the Washington State Department of Health.

[21] The rate of quitting is markedly higher in Washington than Missouri, presumably due in part to several increases in Washington's cigarette tax during the time period. The Missouri tax only had a minor increase from 13 cents per pack to 17 cents per pack in 1994.

Assuming $f = 0.5$ and both ratios $v_{1,1}/v_{0,1}$ and $v_{1,1}/v_{1,0}$ are equal to 3, the fixed effects estimator's probability limit in (20) becomes

$$\text{plim } \hat{\gamma}_{FE} = \gamma \left[\frac{1}{2} \cdot \frac{1}{1+3r} + \frac{1}{2} \cdot \frac{1}{1+3\dfrac{p(1-q)}{(1-p)}} \right] \tag{22}$$

In the case that misclassification is uncorrelated across pregnancies (i.e., $p = q = r$), the expression simplifies to plim $\hat{\gamma}_{FE} = \dfrac{\gamma}{1+3p}$, so that the fixed effects attenuation bias would be $(3+3p)/(1+3p)$ times larger than the OLS attenuation bias. For $p = 0.05$, the OLS and FE biases would be -4.8% and -13.0%, respectively. For $p = 0.10$, the OLS and FE biases would be -9.1% and -23.1%, respectively. If misclassification is correlated across pregnancies (with $r < p < q$), the fixed effects attenuation bias will be reduced. For instance, for $(p, q, r) = (0.05, 0.25, 0.02)$, the FE bias would be equal to -8.1% (as compared to -13.0% in the uncorrelated $p = 0.05$ case). For $(p, q, r) = (0.10, 0.25, 0.02)$, the FE bias would be equal to -12.8% (as compared to -23.1% in the uncorrelated $p = 0.10$ case).

To provide insight on how misclassification might impact the fixed effects results of Section 4, we apply the bias formula in (22) under two different scenarios for the misclassification probabilities. In addition to the parameter values used above (for f and the probability ratios), the probability limit of $\hat{\gamma}_{FE}$ is taken to be -80 (which is slightly conservative based upon the $proxy = 1$ results in Table VI). Figure 4 plots the implied 'true' smoking effect γ as a function of the misclassification probability p. The solid line corresponds to the case of uncorrelated misclassification probabilities ($p = q = r$). Even with an extremely high misclassification rate p of 40%, the implied γ would be equal to -180 g, which is still far lower in magnitude than the OLS estimate. The dashed line in Figure 4 considers a case in which misclassification is correlated across pregnancies ($r < p < q$), with the specific values for q and r given by $q = 2p$ and $p = 2r$ (i.e., twice as likely to misreport given previous misreporting and half as likely to misreport given previous correct reporting). For any given value of p, this case corresponds to less misclassification bias in $\hat{\gamma}_{FE}$.

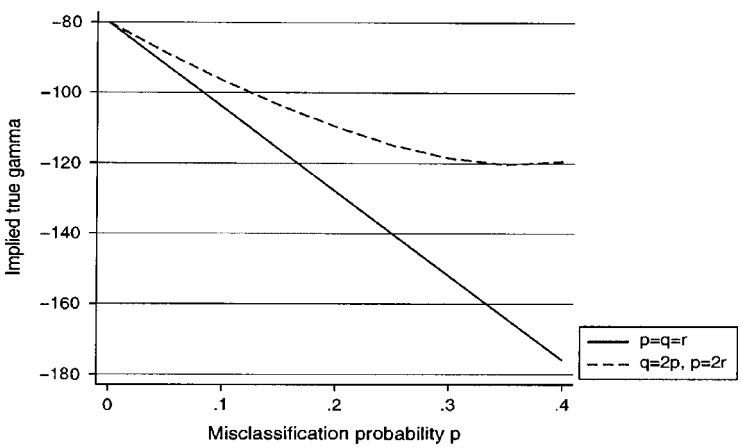

Figure 4. The impact of misclassification on the smoking effect

Although there does not exist direct evidence on misreporting of smoking status by pregnant women, the consistency of smoking rates derived from birth certificate data and other data sources (see Mathews, 2001; Colman *et al*., 2003) suggests that the misclassification probabilities are probably not extremely large. If one assumes that the misclassification probability is in the 10–30% range, the results from Figure 4 suggest that a conservative range for the true effect of smoking on birthweight is a reduction of 100–150 g. Even the upper end of this range (150 g reduction) represents an estimated effect that is 40% lower than the OLS estimated effect.

Finally, we note that the model of misclassification probabilities in (16)–(18) is probably still too simplistic. Although second pregnancy misclassification is allowed to depend upon first pregnancy misclassification, the misclassification probabilities are otherwise assumed to be exogenous. For instance, misclassification of smoking may depend upon the birth outcome or the extent of cigarette usage. As Mathews (2001) points out: '[I]n cases of poor birth outcome, the mother might be less willing to admit having smoked at all during her pregnancy if asked about lifestyle factors after the delivery.' If this relationship exists between the misclassification probability and u, it would likely translate into an upward bias in the magnitude of the fixed effects estimate. The same upward bias could result if women are more likely to misclassify themselves as non-smokers if they are not heavy smokers. In a related context, Kaestner *et al*. (1996) point out this bias for the effect of self-reported drug use upon birthweight. Dependence of misclassification on either birth outcome or extent of smoking could thus serve to offset some of the attenuation bias that results from exogenous misclassification.

6. CONCLUSION

This paper has used matching algorithms to construct panel data sets for estimating the causal effect of smoking on birth outcomes. The fixed effects estimates reported in Section 4 indicate that the effects of smoking on birth outcomes are smaller in magnitude than traditionally estimated by OLS. The lower estimated magnitudes are consistent with the basic omitted variables story where smoking is positively correlated with unobservables that adversely impact birth outcomes.

We have also considered several potential sources of bias for the fixed effects estimates. Three of these potential sources—incorrect matching, feedback effects and correlated changes in maternal behaviour—would likely bias the fixed effects estimator in the direction of OLS (which would imply that the fixed effects estimates are too large in magnitude). However, as discussed in Section 5.3, misclassification of smoking status can cause a bias in the opposite direction. Although there has been no direct evidence on smoking status misclassification in federal natality data, the indirect evidence presented in Section 5.3 suggests that misclassification alone can only account for part of the difference between the fixed effects and OLS estimates reported in Section 4. Given the potential usefulness of panel data in explaining birth outcomes, future follow-up natality surveys might focus efforts on developing a better understanding of smoking status misclassification or requiring cross-validation by physicians.

Based upon the proxy-based fixed effects estimation and the misclassification model, we determined that 100–150 g was a reasonable range for the effect of smoking on birthweight. Looking back at the Surgeon General's statement in the Introduction, the results of this paper reinforce the conclusion that smoking has statistically significant adverse effects on birthweight and is associated with a higher incidence of low-birthweight babies. The actual magnitude of

these effects, however, is probably far lower than traditional OLS estimates would suggest. These reduced effects should be taken into account when considering cost–benefit calculations of smoking-related public policy. In addition, the fact that unobserved behaviour (that is correlated across births) underlies the reduced smoking effect suggests that improving other prenatal care behaviour (proper diet, vitamin supplements, prenatal visits, etc.) remains an extremely important issue for the public health community.

APPENDIX

Kessner Index Details

The Kessner index is a categorical measure for adequacy of prenatal care which is based upon length of gestation, number of prenatal visits and date of initial prenatal visit. The three categories of the Kessner index are 'adequate' (a value of 1), 'intermediate' (a value of 2) and 'inadequate'. Table XI details the criteria for each of the categories. In addition to the specified number of visits required for 'adequate' care in the table, the first prenatal visit must occur during the first 13 weeks (first trimester) in order for care to be considered 'adequate'. Finally, any woman who started prenatal care 28 weeks or later into the pregnancy (third trimester) is categorized as having 'inadequate' care (regardless of the number of visits afterwards).

Additional Tables

Three tables, which are referenced in the main text, are provided in this appendix. Table XII (see Section 2) contains the state pair birth counts for Alabama in 1990. Tables XIII and XIV (see Section 4) report the regression results for the low-birthweight indicator and gestation, respectively.

Table XI. Kessner index for adequacy of prenatal care

Adequacy of care	Gestation (in weeks)	No. of prenatal visits
'Adequate' (value = 1)	≤ 13	1 or more
	14–17	2 or more
	18–21	3 or more
	22–25	4 or more
	26–29	5 or more
	30–31	6 or more
	32–33	7 or more
	34–35	8 or more
	≥ 36	9 or more
'Inadequate' (value = 3)	14–21	0
	22–29	1 or less
	30–31	2 or less
	32–33	3 or less
	≥ 34	4 or less
'Intermediate' (value = 2)	All other combinations not specified above	

Table XII. State pair birth counts for Alabama in 1990

Mother's state of birth	# Births
Alabama	44 652
Alaska	48
Arizona	73
Arkansas	220
California	747
Colorado	113
Connecticut	86
Delaware	36
District of Columbia	83
Florida	1 837
Georgia	2 212
Hawaii	52
Idaho	28
Illinois	1 078
Indiana	482
Iowa	107
Kansas	146
Kentucky	341
Louisiana	601
Maine	58
Maryland	164
Massachusetts	144
Michigan	765
Minnesota	74
Mississippi	1 241
Missouri	291
Montana	33
Nebraska	53
Nevada	21
New Hampshire	18
New Jersey	266
New Mexico	74
New York	776
North Carolina	368
North Dakota	14
Ohio	795
Oklahoma	181
Oregon	40
Pennsylvania	370
Rhode Island	37
South Carolina	294
South Dakota	34
Tennessee	1 158
Texas	885
Utah	48
Vermont	11
Virginia	355
Washington	112
West Virginia	113
Wisconsin	155
Wyoming	18
TOTAL	61 908

Table XIII. Matched panel regressions (dependent variable = low-birthweight indicator)

	Panel #1		Panel #2		Panel #3		Panel #1		Panel #2		Panel #3	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
Smoking	0.0485	0.0308	0.0451	0.0294	0.0429	0.0243	0.0372	0.0195	0.0336	0.0189	0.0309	0.0163
	(0.0019)	(0.0025)	(0.0017)	(0.0021)	(0.0015)	(0.0020)	(0.0035)	(0.0038)	(0.0030)	(0.0032)	(0.0028)	(0.0030)
# cigarettes							0.0010	0.0009	0.0010	0.0009	0.0010	0.0007
							(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Male	−0.0093	−0.0076	−0.0090	−0.0092	−0.0073	−0.0069	−0.0093	−0.0075	−0.0089	−0.0089	−0.0073	−0.0069
	(0.0010)	(0.0013)	(0.0008)	(0.0011)	(0.0007)	(0.0009)	(0.0010)	(0.0013)	(0.0008)	(0.0011)	(0.0007)	(0.0009)
Age	−0.0026	0.0025	−0.0029	0.0001	−0.0028	−0.0007	−0.0029	0.0029	−0.0032	0.0005	−0.0031	−0.0008
	(0.0009)	(0.0023)	(0.0008)	(0.0020)	(0.0007)	(0.0017)	(0.0009)	(0.0023)	(0.0008)	(0.0020)	(0.0007)	(0.0017)
Age^2	0.00006	−0.00004	0.00007	0.00001	0.00006	0.00000	0.00007	−0.00004	0.00007	0.00001	0.00007	0.00000
	(0.00001)	(0.00003)	(0.00001)	(0.00003)	(0.00001)	(0.00002)	(0.00001)	(0.00003)	(0.00001)	(0.00003)	(0.00001)	(0.00002)
High-school graduate	−0.0080	−0.0043	−0.0130		−0.0123		−0.0068	−0.0047	−0.0122		−0.0115	
	(0.0022)	(0.0041)	(0.0020)		(0.0019)		(0.0023)	(0.0042)	(0.0020)		(0.0019)	
Some college	−0.0166	−0.0122	−0.0204		−0.0164		−0.0154	−0.0132	−0.0194		−0.0154	
	(0.0023)	(0.0050)	(0.0020)		(0.0020)		(0.0024)	(0.0051)	(0.0020)		(0.0020)	
College graduate	−0.0223	−0.0132	−0.0246		−0.0217		−0.0208	−0.0150	−0.0235		−0.0206	
	(0.0024)	(0.0064)	(0.0021)		(0.0020)		(0.0024)	(0.0064)	(0.0021)		(0.0020)	
Married	−0.0144	−0.0134	−0.0130	−0.0085	−0.0176		−0.0142	−0.0131	−0.0127	−0.0081	−0.0172	
	(0.0019)	(0.0025)	(0.0016)	(0.0022)	(0.0016)		(0.0019)	(0.0026)	(0.0016)	(0.0022)	(0.0016)	
Black	0.0491		0.0477		0.0479		0.0496		0.0481		0.0484	
	(0.0021)		(0.0020)		(0.0021)		(0.0021)		(0.0020)		(0.0021)	
Kessner index = 2	0.0327	0.0331	0.0293	0.0298	0.0287	0.0279	0.0325	0.0333	0.0289	0.0300	0.0283	0.0278
	(0.0026)	(0.0027)	(0.0022)	(0.0023)	(0.0018)	(0.0019)	(0.0026)	(0.0027)	(0.0022)	(0.0023)	(0.0018)	(0.0019)
Kessner index = 3	0.0890	0.0832	0.0851	0.0752	0.0622	0.0520	0.0884	0.0826	0.0841	0.0756	0.0614	0.0517
	(0.0066)	(0.0057)	(0.0059)	(0.0049)	(0.0052)	(0.0044)	(0.0066)	(0.0057)	(0.0059)	(0.0050)	(0.0052)	(0.0045)
No prenatal visit	0.0160	0.0082	0.0235	0.0104	0.0363	0.0302	0.0156	0.0112	0.0238	0.0111	0.0357	0.0287
	(0.0110)	(0.0089)	(0.0101)	(0.0078)	(0.0093)	(0.0071)	(0.0111)	(0.0090)	(0.0101)	(0.0079)	(0.0093)	(0.0072)
First prenatal visit in 2nd trimester	−0.0319	−0.0310	−0.0274	−0.0272	−0.0254	−0.0226	−0.0318	−0.0309	−0.0273	−0.0274	−0.0252	−0.0226
	(0.0029)	(0.0031)	(0.0025)	(0.0026)	(0.0022)	(0.0023)	(0.0030)	(0.0031)	(0.0025)	(0.0027)	(0.0022)	(0.0023)
First prenatal visit in 3rd trimester	−0.0890	−0.0763	−0.0850	−0.0668	−0.0611	−0.0465	−0.0883	−0.0755	−0.0841	−0.0673	−0.0601	−0.0461
	(0.0074)	(0.0068)	(0.0066)	(0.0059)	(0.0060)	(0.0054)	(0.0075)	(0.0069)	(0.0067)	(0.0060)	(0.0060)	(0.0054)

Table XIV. Matched panel regressions (dependent variable = gestation)

	Panel #1		Panel #2		Panel #3		Panel #1		Panel #2		Panel #3	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
Smoking	−0.188	−0.096	−0.154	−0.056	−0.136	−0.080	−0.154	−0.025	−0.149	−0.015	−0.092	−0.018
	(0.019)	(0.028)	(0.016)	(0.023)	(0.015)	(0.023)	(0.033)	(0.042)	(0.028)	(0.036)	(0.026)	(0.034)
# cigarettes							−0.003	−0.005	−0.001	−0.003	−0.004	−0.005
							(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Male	−0.112	−0.114	−0.120	−0.106	−0.125	−0.106	−0.111	−0.113	−0.120	−0.106	−0.125	−0.107
	(0.011)	(0.014)	(0.009)	(0.012)	(0.008)	(0.010)	(0.011)	(0.014)	(0.009)	(0.012)	(0.008)	(0.010)
Age	0.058	0.069	0.037	0.063	0.043	0.090	0.060	0.066	0.037	0.060	0.043	0.092
	(0.009)	(0.025)	(0.009)	(0.022)	(0.008)	(0.019)	(0.009)	(0.025)	(0.009)	(0.022)	(0.008)	(0.019)
Age^2	−0.0013	−0.0005	−0.0010	−0.0005	−0.0010	−0.0009	−0.0013	−0.0005	−0.0010	−0.0005	−0.0010	−0.0009
	(0.0002)	(0.0003)	(0.0001)	(0.0003)	(0.0001)	(0.0003)	(0.0002)	(0.0003)	(0.0001)	(0.0003)	(0.0001)	(0.0003)
High-school graduate	0.012	−0.002	0.032		0.032		0.009	0.005	0.030		0.029	
	(0.023)	(0.045)	(0.020)		(0.020)		(0.024)	(0.046)	(0.020)		(0.020)	
Some college	0.024	0.000	0.061		0.035		0.021	0.013	0.059		0.031	
	(0.025)	(0.055)	(0.021)		(0.021)		(0.025)	(0.056)	(0.021)		(0.021)	
College graduate	0.056	0.027	0.088		0.072		0.051	0.042	0.086		0.068	
	(0.026)	(0.070)	(0.022)		(0.022)		(0.026)	(0.071)	(0.022)		(0.022)	
Married	0.132	0.053	0.113	0.048	0.159		0.129	0.050	0.110	0.045	0.159	
	(0.020)	(0.028)	(0.017)	(0.024)	(0.017)		(0.020)	(0.028)	(0.017)	(0.024)	(0.017)	
Black	−0.571		−0.616		−0.604		−0.569		−0.615		−0.604	
	(0.021)		(0.020)		(0.021)		(0.021)		(0.021)		(0.021)	
Kessner index = 2	−0.356	−0.309	−0.378	−0.319	−0.404	−0.341	−0.359	−0.311	−0.378	−0.323	−0.403	−0.340
	(0.024)	(0.030)	(0.020)	(0.025)	(0.017)	(0.021)	(0.024)	(0.030)	(0.020)	(0.025)	(0.017)	(0.021)
Kessner index = 3	−0.803	−0.571	−0.762	−0.592	−0.686	−0.504	−0.815	−0.596	−0.757	−0.591	−0.690	−0.518
	(0.056)	(0.062)	(0.051)	(0.054)	(0.047)	(0.049)	(0.057)	(0.063)	(0.051)	(0.055)	(0.047)	(0.050)
No prenatal visit	−0.219	−0.360	−0.348	−0.402	−0.290	−0.476	−0.194	−0.338	−0.335	−0.406	−0.262	−0.453
	(0.106)	(0.097)	(0.098)	(0.086)	(0.092)	(0.079)	(0.107)	(0.098)	(0.099)	(0.087)	(0.093)	(0.080)
First prenatal visit in 2nd trimester	0.492	0.401	0.520	0.426	0.533	0.401	0.497	0.404	0.518	0.428	0.530	0.398
	(0.028)	(0.034)	(0.025)	(0.029)	(0.022)	(0.025)	(0.029)	(0.034)	(0.025)	(0.029)	(0.022)	(0.025)
First prenatal visit in 3rd trimester	0.981	0.779	0.926	0.734	0.834	0.587	0.984	0.795	0.910	0.721	0.826	0.586
	(0.067)	(0.075)	(0.060)	(0.065)	(0.056)	(0.060)	(0.068)	(0.076)	(0.061)	(0.066)	(0.057)	(0.060)

ACKNOWLEDGEMENT

Seminar participants at Columbia, Purdue, Illinois and the 11th International Conference on Panel Data provided helpful comments. The author is particularly grateful to Herman van Dijk (co-editor) and two anonymous referees for suggestions that substantially improved this paper. Javaria Ahmad provided excellent research assistance. Financial support was provided by the John & Mary Willis Young Faculty Scholar fund at Purdue University and the faculty research fund at the University of Chicago GSB.

REFERENCES

Colman G, Grossman M, Joyce T. 2003. The effect of cigarette excise taxes on smoking before, during, and after pregnancy. *Journal of Health Economics* **22**: 1053–1072.

Corman H. 1995. The effects of low birthweight and other medical risk factors on resource utilization in the pre-school years. NBER Working Paper 5273.

Corman H, Chaikind S. 1998. The effect of low birthweight on the school performance and behavior of school-aged children. *Economics of Education Review* **17**: 307–316.

Currie J, Gruber J. 1996. Saving babies: the efficacy and cost of recent changes in the Medicaid eligibility of pregnant women. *Journal of Political Economy* **104**: 457–470.

Currie J, Hyson R. 1999. Is the impact of health shocks cushioned by socioeconomic status? The case of low birthweight. *American Economic Review* **89**: 245–250.

Currie J, Moretti E. 2002. Mother's education and the intergenerational transmission of human capital: evidence from college openings and longitudinal data. NBER Working Paper 9360.

Evans WN, Lien DS. 2005. The benefits of prenatal care: evidence from the PAT bus strike. *Journal of Econometrics* **125**: 207–239.

Evans WN, Ringel JS. 1999. Can higher cigarette taxes improve birth outcomes? *Journal of Public Economics* **72**: 135–154.

Freeman RB. 1984. Longitudinal analyses of the effects of trade unions. *Journal of Labor Economics* **2**: 1–26.

Griliches Z, Hausman JA. 1986. Errors in variables in panel data. *Journal of Econometrics* **31**: 93–118.

Hack M, Klein NK, Taylor HG. 1995. Long-term developmental outcomes of low birth weight babies. *The Future of Children* **5**: 19–34.

Jakubson G. 1986. Measurement error in binary explanatory variables in panel data models: why do cross section and panel estimates of the union wage effect differ? Princeton University, Industrial Relations Section Working Paper #209.

Joyce T. 1999. Impact of augmented prenatal care on birth outcomes of Medicaid recipients in New York City. *Journal of Health Economics* **18**: 31–67.

Kaestner R, Joyce T, Wehbeh H. 1996. The effect of maternal drug use on birth weight: measurement error in binary variables. *Economic Inquiry* **34**: 617–629.

Levinson A, Ullman F. 1998. Medicaid managed care and infant health. *Journal of Health Economics* **17**: 351–368.

Lewit EM, Baker LS, Corman H, Shiono PH. 1995. The direct costs of low birth weight. *The Future of Children* **5**: 35–51.

Lien DS, Evans WN. 2001. Estimating the impact of large cigarette tax hikes: the case of maternal smoking and infant birth weight. Mimeo.

Mathews TJ. 2001. Smoking during pregnancy in the 1990's. National Vital Statistics Reports 49(7), Centers for Disease Control and Prevention.

Peltzman S. 1975. The effects of automobile safety regulation. *Journal of Political Economy* **83**: 677–725.

Permutt T, Hebel JR. 1989. Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight. *Biometrics* **45**: 619–622.

Reichman NE, Florio MJ. 1996. The effects of enriched prenatal care services on Medicaid birth outcomes in New Jersey. *Journal of Health Economics* **15**: 455–476.

Rosenzweig MR, Wolpin KI. 1991. Inequality at birth: the scope for policy intervention. *Journal of Econometrics* **50**: 205–228.
Schoendorf KC, Kiely JL. 1992. Relationship of sudden infant death syndrome to maternal smoking during pregnancy. *Pediatrics* **90**: 905–908.
Walsh RA. 1994. Effects of maternal smoking on adverse pregnancy outcomes: examination of the criteria of causation. *Human Biology* **66**: 1059–1062.
Wooldridge JM. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.

