Maternal and Child Health Journal, Vol. 4, No. 1, 2000

Methodological Note

Measurement Error in Prenatal Care Utilization: Evidence of Attenuation Bias in the Estimation of Impact on Birth Weight

John R. Penrod, PhD,[1,3] and Paula M. Lantz, PhD[2]

Objective: Errors in the measurement of the timing and number of prenatal care visits may produce downward bias in estimates of the impact of prenatal care use on birth outcomes. This paper examines the extent of attenuation bias from measurement error in the estimation of the effect of prenatal care use on birth weight. *Methods:* Data were analyzed from the 1980 National Natality Survey, a nationally representative sample of live births with information on prenatal care utilization from three sources: birth certificates, medical provider surveys, and maternal surveys. The extent of attenuation bias in estimates of the impact of different measures of prenatal care use on birth weight was examined by comparing estimates robust to measurement error (including instrumental variables) with ordinary least squares results. *Results:* There is considerable disagreement in measures of prenatal care across the three data sources, with correlations in the utilization measures computed from different sources around 0.5. The results also show evidence of attenuation bias from measurement error in estimates of the impact of prenatal care on birth weight for both White and Black mothers. Attenuation bias was least severe for information from the birth certificate report of prenatal care. *Conclusions:* Because of measurement error, previous studies may have underestimated the effect of prenatal care utilization on birth weight. Corrected estimates, however, do not suggest that prenatal care is a major predictor of birth weight. In addition, part of what previous analyses have interpreted as adverse selection bias may in fact be attenuation bias due to measurement error.

KEY WORDS: Prenatal care utilization; low birth weight; selection bias; attenuation bias; health policy.

INTRODUCTION

Low birth weight (<2500 g) and its major antecedent of prematurity (<37 completed weeks of gestation) are serious public health problems in the United States (1, 2). To date, the primary focus of policies and interventions attempting to alleviate the problems of low birth weight and prematurity has been on prenatal care (3). For example, during the 1980s, Congress enacted a series of Medicaid eligibility expansions to increase the early and continued use of prenatal care among economically disadvantaged women. In addition, many health advocates continue to call for universal maternity care as a way to reduce significantly the public health burden of low birth weight and infant mortality (1, 4).

Despite the strong policy focus on prenatal care, the research literature on the relationship between prenatal care and birth weight remains inconclusive (5–8). As with many medical interventions, ethical considerations have prevented a randomized controlled trial to establish the efficacy of standard pre-

[1]Faculty of Medicine, McGill University/The Montreal General Hospital, Montreal, QC, Canada.
[2]Department of Health Management and Policy, University of Michigan School of Public Health, Ann Arbor, Michigan.
[3]Correspondence should be directed to John Penrod, PhD, Centre for the Analysis of Cost-Effective Care, The Montreal General Hospital, 1650 Cedar Avenue, Montreal, QC, Canada H3G 1A4; e-mail: mdpn@musica.mcgill.ca

natal care. Therefore, analysts have had to rely on patterns of birth weight, prenatal care use, sociodemographic characteristics, and the availability of health care and health insurance in the population to identify effects. While prenatal care utilization is significantly related to birth weight in simple bivariate analyses, it actually has not emerged as a strong predictor of birth weight in the majority of published multivariate analyses (5, 9–11). Furthermore, the results of exercises estimating the maximum effect of adequate prenatal care utilization suggest that prenatal care is not a panacea for improving infant health at the population level or for reducing racial disparities in birth outcomes (12–14).

Serious methodological problems continue to impede research on the impact of prenatal care utilization on birth weight and other birth outcomes (15, 16). One such issue that has been the subject of considerable discussion in the literature is selection bias: differences in the ways in which women seek and receive prenatal care may be related to unmeasurable and/or unobserved risks for poor birth outcomes, biasing estimates of the impact of prenatal care (8, 17). Researchers generally have found evidence for ‘‘adverse selection’’—that is, women at higher risk for poor birth outcomes tend to receive prenatal care earlier and more frequently, resulting in an underestimation of the effect of prenatal care.

A second methodological issue, which is the focus of this paper, is error in the measurement of prenatal care utilization and resulting bias in estimates of birth weight. Although a variety of data sources are used in studies of birth weight predictors, one frequently used source is birth registration data. The reliability and validity of many of the data items on the U.S. Standard Certificate of Live Birth have been studied and shown to be of relatively high quality. However, the quality of birth certificate information on prenatal care utilization is likely to have some measurement error (18–20). It is possible that birth certificate data do not accurately reflect the actual amount and timing of prenatal care received, with perhaps the greatest discrepancies for those groups at risk for low birth weight. Thus, errors in the measurement of women’s prenatal care utilization may lead to bias in estimates of the impact of prenatal care on birth outcomes. Since the most likely effect of any bias here is an attenuation effect, the result could be an underestimation of the impact of prenatal care on birth weight.

Previous studies suggest that there are indeed inaccuracies in reports of prenatal care timing and adequacy across a variety of data sources, including birth certificate data (21–23). Also, Kogan *et al.* (24) suggest that different approaches to the measurement of prenatal care produce different results regarding utilization rates, thus making it difficult to interpret previous trend analyses. The subject of measurement attenuation bias in estimates of prenatal care efficacy, however, has received limited empirical attention in the research literature to date.

The goal of our research was to investigate the existence and consequences of errors in the measurement of prenatal care utilization using a nationally representative dataset that includes information from birth certificates, medical care provider records, and a maternal questionnaire. Our study aims included the following: (1) to determine the level of agreement among three important sources of information on prenatal care use (birth certificate data, medical provider reports, and maternal self-reports) using a nationally representative sample; (2) to identify sociodemographic characteristics related to discrepancies across the three data sources; and (3) to determine whether error in the measurement of prenatal care use has a biasing effect on estimates of the association between prenatal care use and birth weight. We accomplish this by comparing estimates from analyses that correct for measurement error with estimates from analyses that do not correct for measurement error.

The role of prenatal care in public health policies aimed at reducing negative birth outcomes will remain unclear in the absence of improvements in its measurement (25). While the results of several studies suggest that standard prenatal care is not a major determinant of birth outcomes such as birth weight and gestational age, difficulties in the measurement of prenatal care mean that prior studies may have underestimated its potential. The research and results described here offer guidance for interpreting past research and for conducting future research on the relationship between prenatal care and birth outcomes such as low birth weight. Such work is critical to understanding the relationship between prenatal care use and birth outcomes, and for advancing public health policy aimed at improving maternal and infant health.

METHODS

Data

This study used data from the 1980 National Natality Survey (NNS), a follow-back survey designed to provide detailed information on the pregnancies, labors, and deliveries of a nationally representative sample of live births. A probability sample of 9941 live births was selected for this survey, with low-birth-weight infants being oversampled (26). The information in the NNS came from several sources, including birth certificates, medical records information (from questionnaires mailed to hospitals, attendants at delivery, prenatal care providers, and providers of radiation examinations and treatments), and questionnaires mailed to mothers who were married at the time of delivery, providing three independent sources of information about the same basic elements of prenatal care utilization. The 1980 NNS remains a unique and important resource because of its scope and its three separate data sources.

Information on several aspects of prenatal care utilization are available in the birth certificate, medical record, and maternal questionnaire segments of the NNS, including the month of pregnancy in which prenatal care began (which is viewed as the "delay" between conception and the initiation of care) and the total number of prenatal care visits (truncated at 30 visits in the medical records portion). These two frequently used measures of prenatal care utilization, along with data on gestational age at delivery, also provide the basis from which two well-known categorizations of prenatal care adequacy are constructed. The Kessner Index categorizes prenatal care use as being either "inadequate," "intermediate," or "adequate" (15, 27, 28). The Adequacy of Prenatal Care Utilization (APNCU) Index categorizes the care received as being "inadequate," "intermediate," "adequate," or "adequate plus" (29, 30). Women receive "adequate-plus" care primarily because of pregnancy complications and poor maternal health status, situations which place these women at high risk for a negative reproductive outcome.

There is little published work describing information consistency across the NNS data sources. Fingerhut and Kleinman (23) compared NNS data on the trimester of prenatal care utilization from birth certificates to information from the mother's questionnaire (they omitted comparisons to the medical record data), showing significant discrepancies. They found that mothers tended to report that care began earlier than was indicated by the birth certificate. However, their study was limited to a comparison of the data on trimester of initiation; they did not compare data on items such as exact month of care initiation or total number of visits, items which form the basis of prenatal care utilization indices.

Forrest and Singh (22) compared individual-level NNS data from birth certificates to data from the medical records regarding timing of prenatal care initiation, finding agreement in the month that prenatal care began in only 35% of cases. In 56% of cases, the medical record information suggested that care began later than indicated on the birth certificate. In addition, they found that the rate of discrepancy in the reports of the month of care initiation varied with sociodemographic characteristics, although they also found that even the "best" subgroups had high proportions with discrepant information. They conclude that "the differences among data sources are more fundamental and important than any differences in the quality of reporting among socioeconomic subgroups" (22).

None of these three data sources can be regarded as a "gold standard." However, as is shown below, we are able to learn a good deal more from multiple noisy measures of the same variable than if we just had one noisy measure. The advantage of the method we employ—instrumental variables—is that none of the measures need be the gold standard in order to produce a consistent estimate of the association between prenatal care utilization and birth weight.

Study Design and Methodology

Aim 1: Analysis of Utilization Discrepancies

Since the maternal questionnaire was not sent to unmarried women in the NNS probability sample (approximately one of five cases), comparisons between birth certificate and medical record reports can be made for the whole sample, but comparisons using maternal self-report data can be made only with the sample of married women. Therefore, we constructed two separate samples for analysis. Sample 1 included all live births for which there was a response to the hospital/provider survey that included medical records information on prenatal care utilization, omitting cases for which prenatal care information in the birth certificate file was missing or imputed from other NNS sources. The resulting sample included 8214 live births (82.8% of the entire sample). Sample 2 included married women from Sample 1 who responded to the maternal questionnaire and had non-imputed data on prenatal care utilization in the maternal questionnaire (over 40% of women who responded to the maternal questionnaire had imputed prenatal care utilization data). Sample 2 included 3924 live births (39.5% of the overall sample).

Our descriptive work built on prior analyses of NNS data on prenatal care utilization in two ways. First, we extended the description of data source discrepancies regarding prenatal care use to items beyond the month of care initiation. In our study, we examined four measures of prenatal care utilization, including the month of care initiation or delay in care initiation (coded as months 1–10, with 10 being no prenatal care), the total number of visits made (ranging from 0 to 30 or more), the Kessner Index score (1 = inadequate, 2 = intermediate, 3 = adequate), and the APNCU Index score (1 = inadequate, 2 = intermediate, 3 = adequate, and 4 = adequate plus).

Aim 2: Identifying Characteristics Associated with Discrepancies

We also investigated sociodemographic characteristics related to discrepancies across NNS data sources. While prior work was limited to identifying characteristics related to discrepancies in timing of care initiation, we identified characteristics related to discrepancies regarding other components of prenatal care utilization as well. We first calculated the degree and direction of differences between data sources for the four components of prenatal care utilization under study and ran OLS regressions to identify sociodemographic characteristics associated with increased levels of misclassification across sources for each of the four measures of care. The index of discrepancy (or dependent variable) we employed was the squared difference in the measure of prenatal care across the two sources.

Independent variables taken from the birth certificate file included age of mother, race of mother, education of mother and education of father (measured as years of school completed), marital status of mother, residence in a metropolitan county, plurality (multiple birth versus not), parity (or total birth order), and previous termination (the previous pregnancy ended in termination). Measures of the outcome of the pregnancy were also included as independent variables, under the hypothesis that a negative outcome may induce a bias in one or more of the reports of prenatal use, perhaps impacting the rate of matching across data sources. It is possible, for example, that mothers who delay care and have poor birth outcomes systematically understate the extent of the delay. Outcome variables studied as determinants of the discrepancy include gestational age (measured in weeks), birth weight (measured in grams), and congenital malformations (malformation present versus none).

Aim 3: Assessing Attenuation Bias

To achieve the third aim of this study, we described the effect of measurement error in two types of models. In the first type, prenatal care utilization was described as a continuous variable (i.e., number of visits, delay in treatment) and in the second type, prenatal care adequacy was described by a categorical index set (i.e., Kessner Index and APNCU Index). In the former case, the extent of attenuation bias was evaluated by comparing OLS estimates to two-stage least squares (TSLS) estimates, which use the redundant data on prenatal care utilization to correct for measurement error. In the latter case, OLS estimates with prenatal care variables constructed from one data source were compared to results estimated from variables using data from multiple sources.

To illustrate the methodology underlying this component of our study, suppose a quantity of interest Z is not observed, but that we have two measures, Z_1 and Z_2, that both measure Z with some error. Employing ideas similar to Ashenfelter and Krueger (31), we describe below how measurement error leads to attenuation bias in OLS regressions and how instrumental variable estimates can be used to obtain corrected estimates. A more detailed mathematical appendix deriving these results is available from the authors.

Assume for now that the measurement error is classical: that is, the measures are unbiased; the errors are uncorrelated with the truth (Z) or with each other; and the errors have equal variance. More generally, however, one measure may be noisier than another or the two measures may be correlated. Suppose that one is interested in estimating a linear regression model in which Z is hypothesized to affect outcome BW:

$$BW = \alpha + Z\beta + u$$

It is well known that if one of the noisy measures of Z is used in the estimation of such an equation, then β_{OLS}, the OLS estimator of the slope coefficient β, is biased toward zero. For example, if Z_1 is used in estimation, the probability limit of the OLS estimator is biased downward according to a ratio that is sometimes referred to as an attenuation coefficient or as the "signal to total variance" ratio for the measure Z_1. For the case of classical measurement error in the two measures, it is easily shown that this attenuation factor is equal to the correlation coefficient of the two measures. For example, if the correlation coefficient of the two measures is 0.5, other things equal, this suggests that the true effect of prenatal care on birth weight outcomes may be attenuated by a factor of 50% when this effect is estimated using a single noisy measure.

It is possible to obtain a consistent estimate of the coefficient β by the method of instrumental variables. Instrumental variables (IV) is a method frequently used in many econometric applications, including measurement error (31). When two alternative measures of a quantity are available, IV uses information from the covariance between the two measures to construct an unbiased estimate of the coefficient of interest. The extent of attenuation bias can then be determined by comparing the IV estimates of β to the estimates from OLS regressions that use noisy measures of Z.

The analysis in the multiple regression context is similar to the bivariate case; here we are simply regression-adjusting the birth weight variable. We estimated regressions of the following type:

$$BW_i = \alpha + Z_i\beta + H_i\delta + X_i\gamma + \varepsilon_i$$

where BW_i is the birth weight of infant i, Z_i is a measure of prenatal care utilization, H_i is a set of health and fertility variables shown to be associated with birth weight, and X_i is a set of demographic variables of the mother and father related to birth weight. In our model, H_i included the following variables: mother's age and its square, birth order, plurality, and whether there was a previous terminated pregnancy. X_i included dummy variables for levels of mother's and father's education and an indicator for metropolitan residence. Since Joyce (32) and others have shown that birth weight functions for Blacks and Whites are statistically different, we estimated separate models for Blacks and Whites. Small numbers precluded us from doing any further subgroup analysis by race or ethnicity.

In the multiple regression case, we employed the method referred to as two-stage least squares (TSLS). In the first stage of the estimation, one measure of prenatal care utilization was regressed on the full set of variables in the birth weight equation plus the alternative measure of prenatal care utilization. In the second stage, birth weight was regressed on the predicted value of prenatal care utilization from the first stage along with the other variables of interest. The standard errors were adjusted accordingly, since the prenatal care utilization variable in the second stage was based on a predicted rather than observed value. The two-stage procedure is implemented as an automated routine in many statistical packages. As in the bivariate case, in multiple regression the extent of bias in the OLS estimate of β was established by comparing estimates from OLS and TSLS estimation.

Up to this point we have discussed the case of classical measurement error. An important assumption of the model and a condition for the IV estimates to be consistent is that the instrument must be uncorrelated with the error term in the birth weight equation. However, suppose that in the retrospective maternal reports, mothers with lower birth weight infants tend to overstate their utilization of prenatal care while mothers with better outcomes understate prenatal care utilization, but that the measurement error in the birth certificate report remains classical. In this case, the estimate where the maternal report was instrumented with the birth certificate report will remain consistent, correcting for both the attenuation bias due to the random component of the error as well as the bias introduced by the systematic component. However, the estimate where the birth certificate report was instrumented with the maternal self report is not consistent. We will turn to this possibility in the empirical work.

Correction of Measurement Error in Categorical Variables

Measurement error in the underlying data on timing and number of prenatal care visits will result in some misclassifications of the Kessner and APNCU indices. However, instrumenting a set of dummy variables measured with error with another set of dummy variables, also measured with error, does not yield a consistent estimate of the effect of prenatal care, since the measurement error in both sets of dummies departs from classical assumptions. Therefore, we implemented a strategy following Card (33), whose work suggests that persons who are classified the same way in two or more data sources are more likely to actually be in that category than those whose classification varies across sources. Thus, estimates for variables based on categorization matches across data sources will have less attenuation bias. We implemented this idea by constructing a set of dummy variables based on comparisons of each person's scores on both the Kessner and the APNCU indices from the birth certificate and medical record. For example, we assigned the variable K_{32} equal to one if the Kessner index category on the birth certificate was adequate (coded as 3) while the Kessner classification in the medical record was intermediate (coded as 2). We estimated multivariate models predicting birth weight with these sets of dummy variables, omitting the category K_{11} (inadequate care on both reports). Evidence of attenuation bias occurs if the effect of prenatal care use on birth weight is stronger for cases where the utilization index category matches across data sources than for cases where it does not.

RESULTS

Descriptive statistics on prenatal care utilization measures and other key variables from the NNS data sources are shown by race in Table I, separately for Sample 1 (all women with nonimputed prenatal care information on the birth certificate and medical records segments) and Sample 2 (those married women with nonimputed information on all three segments). Birth weight and gestational age were fairly consistent across the two NNS data sources for both Whites and Blacks, as shown by the similarity in the mean values and the high correlation coefficients. Information on prenatal care utilization, however, was less consistent, with correlations ranging from .42 to .60. In Sample 1, the mean timing of care initiation was approximately a full month earlier in the birth certificate file than in the medical records file for both races, and the mean number of prenatal care visits was slightly higher in the birth certificate data than the medical records data. Given the discrepancies in important components of prenatal care use, it is not surprising that the correlations between the two data sources in regard to the Kessner Index and APNCU Index—which are based on these components—were also low for both White and Black mothers. Also note that, according to data from both the birth certificate and medical records, the average birth weight, gestational age, and use of prenatal care was lower for Black mothers than White mothers. Similar patterns were observed by race in Sample 2.

Table II presents additional results comparing prenatal care utilization components across NNS data sources for Sample 1, providing an indication of the magnitude and direction of the discrepancies. For month of care initiation, only 36.2% of Whites and 27.5% of Blacks had matching information, with the majority of the discrepancy stemming from birth certificate reports of earlier care initiation. For the total number of prenatal care visits, 43.7% of Whites and 41.6% of Blacks either matched across data sources or were within two visits of matching. For White

Table I. Descriptive Statistics from the 1980 National Natality Survey

A. Means and correlations of prenatal care utilization measures across data sources by race for Sample 1

	Whites			Blacks		
Variable	BC	MR	*r*	BC	MR	*r*
Birth weight (g)	3210.3	3214.9	.99*	2776.1	2782.9	.98*
Gestational age (weeks)	39.0	38.8	.80*	37.6	37.5	.84*
Month prenatal care began	2.79	3.58	.60*	3.35	4.38	.55*
Number of prenatal visits	10.6	10.1	.45*	8.8	8.11	.47*
Kessner Index	2.62	2.41	.47*	2.33	2.05	.46*
APNCU Index	2.73	2.59	.42*	2.49	2.17	.43*
Sample size		$n = 6{,}724$			$n = 1{,}249$	

B. Mean values comparing prenatal care utilization measures across data sources by race for Sample 2

Variable	BC	MR	MQ	BC	MR	MQ
Birth weight (g)	3265.7	3267.6	na	2868.2	2872.35	na
Gestational age (weeks)	39.1	39.0	na	37.6	37.6	na
Month prenatal care began	2.52	3.22	2.22	2.83	3.81	2.65
Number of prenatal visits	11.0	10.7	12.2	9.4	8.7	11.1
Kessner Index	2.72	2.54	2.82	2.50	2.18	2.57
APNCU Index	2.85	2.77	3.07	2.65	2.48	2.95
Sample size		$n = 3{,}623$			$n = 196$	

Note: BC = birth certificate; MR = medical record; MQ = medical questionnaire.
*$p < .01$.

Table II. Distribution of Matching of Prenatal Care Utilization Measures Across Data Sources in the National Natality Survey by Race (Sample 1)

Prenatal care utilization measure	White, % (N = 6,724)	Black, % (N = 1,249)
Birth certificate reports month of care initiation		
2 or more months earlier than medical record	20.5	31.6
1 month earlier than medical record	34.4	29.4
Same as medical record	36.2	27.5
1 month later than medical record	6.3	7.3
2 or more months later than medical record	2.6	4.2
Total	100	100
Birth certificate reports number of visits		
3 or more visits greater than medical record	24.9	23.1
Same as medical record within 2 visits	43.7	41.6
3 or more visits less than medical record	31.5	35.3
Total	100	100
Birth certificate Kessner Index		
Same as medical record	64.4	55.9
1 category different than medical record	31.2	37.2
2 categories different than medical record	4.4	6.9
Total	100	100
Birth certificate APNCU Index		
Same as medical record	46.4	46.2
1 category different than medical record	36.2	29.0
2 categories different than medical record	13.2	16.8
3 categories different than medical record	4.2	8.0
Total	100	100

women, the birth certificate reported three or more visits than the medical record in 31.5% of the cases and underreported by three or more visits in 24.9% of the cases. The pattern was similar for Black women. With regard to the prenatal care utilization indices, 64.4% of Whites and 55.9% of Blacks had matching Kessner Index scores, while 46.4% of Whites and 55.9% of Blacks had matching APNCU index scores.

We repeated the above comparison for Sample 2 (results not shown), and again found that there is a great deal of disagreement across the NNS data sources for both White and Black births, with the degree of matching lower for Black cases in the majority of the comparisons. The lowest degree of matching was between the medical record and the maternal questionnaire for all four prenatal care variables. For example, in terms of the month of care initiation, 43.9% of birth certificate reports for White women and 32.1% of reports for Black women corresponded with the maternal self-reports. In terms of the Kessner Index, 77.0% of the birth certificate and maternal reports matched for White women and 56.6% matched for Black women.

Ordinary least squares regression analysis (using Sample 1) was conducted to see if any characteristics of the mother or the birth outcome were associated with a higher degree of discrepancy between the medical record and birth certificate reports of prenatal care use (as measured by the squared difference between the medical record and birth certificate reports of prenatal care measures). Several maternal characteristics, including younger age, being Black, and metropolitan county of residence, were significantly associated with discrepancies for all measures of prenatal care utilization (Table III). Many of the characteristics associated with data source discrepancies in prenatal care use variables are also known risk factors for low birth weight. These results suggest that those women at higher risk for low birth weight were also more likely to have a lack of agreement regarding prenatal care measures across the data sources under study, meaning that there might be some systematic as well as random error in the measurement of prenatal care. However, because of its administrative nature, we assume that the birth certificate report of prenatal care is the least likely of the three reports to have any systematic error in its reporting (i.e., the

Table III. Coefficients from OLS Regressions Predicting Squared Difference Between Medical Record and Birth Certificate Sources of Prenatal Care Information (Sample 1, $n = 8{,}214$)

Independent variables	Month of initiation	Number of visits	Kessner Index	APNCU Index
Intercept	1.1	23.4**	0.65**	4.32**
Maternal age	−0.07**	0.07	−0.009*	−0.018*
Maternal race Black	1.04**	1.65	0.10**	0.25**
Maternal race other	0.65	−0.79	0.08	−0.07
Maternal education	−0.10*	−0.40	−0.02*	−0.01
Father's education	−0.02	−0.22	−0.001	−0.027
Married	−1.28**	−0.64	−0.06*	−0.20*
Metropolitan county	0.46**	4.95**	0.08**	0.15*
Multiple birth	0.18	0.90	−0.03	0.06
Parity	0.25**	0.26	0.03*	0.04
Previous termination	−0.19	1.93	−0.07*	−0.14
Birth weight	−0.0006**	−0.0001	−0.00003*	−0.00003
Gestational age	0.24**	−0.02	0.01**	−0.067**
Congenital malformation	0.38	−1.21	0.04	0.09

$^*p < .05$; $^{**}p < .01$.

degree of error in prenatal care measurement is not significantly correlated with birth weight or other variables). If there is no systematic bias in prenatal care measurement on the birth certificate, using this report as an instrument in TSLS analyses would yield consistent estimates of the impact of prenatal care on birth weight, and it is these estimates that we discuss in the section below.

Birth Weight Estimation

In Table IV, we present birth weight estimates, using the number of visits as the measure of prenatal care utilization and controlling for gestational age (in weeks) at delivery, maternal education, paternal education, maternal age and its square, total birth order, previous pregnancy termination, and metropolitan residence. For White women in Sample 1, OLS estimates from both birth certificate (BC) and medical record (MR) data indicated that an additional prenatal care visit yielded an additional 14–15 g of birth weight (or about 1/2 ounce). In the final column we present TSLS results where the birth certificate report served as an instrument for the medical record report. We can see that the TSLS estimate is more than double those of the OLS specification, with an increase in birth weight of 38 g per visit. The

Table IV. Grams of Birth Weight Associated with Number of Prenatal Care Visits by Race

Sample 1		OLS BC	OLS MR	TSLS MR-BC
Birth weight per visit (g)				
Whites		15.1**	14.3**	38.1**,+
Blacks		15.6*	5.65	38.5**,+

Sample 2	OLS BC	OLS MR	OLS MQ	TSLS MR-BC	TSLS MQ-BC
Birth weight per visit (g)					
Whites	14.9**	16.7**	8.57**	38.5**,+	45.0**,+
Blacks	16.9*	6.38	−8.88	41.9	56.5

Note: The label BC refers to the birth certificate report of prenatal care visits, while MR and MQ refer to the medical record and mothers' questionnaire reports, respectively. In the TSLS estimation, the heading MR-BC refers to the MR report instrumented with the BC report. The regression analysis also included gestational age, three indicators for mother's level of education, three indicators for father's education, mother's age and its square, total birth order, an indicator for a previous termination, and metropolitan residence.
$^*p < .05$; $^{**}p < .01$; $^+p < .05$ for Durbin–Wu–Hausman test, comparing TSLS result from OLS result from birth certificate.

Durbin–Wu–Hausman test indicated a significant difference between the OLS and TSLS estimate (27).

The patterns in the results for the sample of Black mothers (Table IV) were similar to those for White mothers. The results for both White and Black women from Sample 1 suggested that estimates of the effect of the number of visits on birth weight corrected for measurement error were significantly greater than those from OLS models. A prenatal care visit appeared to have the effect of 38 g (or 1.3 ounces) of birth weight for both White and Black mothers. Although statistically significant, the clinical significance of this level of increase in birth weight is not clear. For example, increasing the number of prenatal care visits by three would lead to only 3.9 additional ounces of birth weight.

For Sample 2, three sources of prenatal care data made more TSLS estimates possible. As shown in Table IV, the effect of prenatal care visits from the birth certificate and the medical record were similar to those found in Sample 1. The results also provided evidence of attenuation bias in the measure of prenatal care visits, since the TSLS estimates of the impact of an additional prenatal care visit were significantly greater than their OLS counterparts for White mothers.

When using month of care initiation (or delay) as the measure of prenatal care utilization, we found that—for White mothers using OLS analysis—the birth certificate estimate was significantly related to birth weight, while the medical record estimate was not (Table V, Sample 1). The TSLS estimate for White women indicated a statistically significant 17.2-g reduction in birth weight for each month of delay, which was statistically different from the 10.2-g reduction estimated in OLS analysis using birth certificate data. The results from Sample 2 for White women were similar.

The results regarding delay in care initiation for Black women (Table V) also showed that the birth certificate report was significantly associated with birth weight, while the medical record was not. The TSLS results for Sample 1 Black women suggested that the TSLS estimate (−36.8 g) for each month of care delay was significantly different from its OLS equivalent (−19.9 g). In addition, the estimated benefit of 1-month reduction in delay was larger for Black mothers than for White mothers (36.8 g versus 17.2 g). Small numbers preclude precise estimates for Black mothers from Sample 2.

After investigating the delay and number of visits measures separately, we investigated the frequently used categorizations of prenatal care adequacy—the Kessner Index and the APNCU Index. Results for the Kessner Index suggested that it was more strongly associated with birth weight using data from birth certificates than data from medical records (Table VI). Relative to receiving inadequate care, OLS results for both White mothers (column 2) and Black mothers (column 5) indicated a statistically significant effect of both intermediate and adequate prenatal care utilization on birth weight using birth certificate data. The effects of both intermediate and adequate care were greater for Black women.

The results for the Kessner Index (Table VI) suggested that there was not much attenuation in OLS estimates of the impact of intermediate care on birth weight for White women (column 4) or Black women (column 7) when based on the birth certificate data, while the OLS estimates based on medical record data were severely attenuated. For example, for White women, the OLS estimate for the effect of intermediate care (relative to inadequate care) was 48.5 g using data from the birth certificate, 9.55 g using data from the medical record, and 45.4 g in the analysis for mothers whose Kessner Index score was the same in both the birth certificate and medical record files. However, both OLS estimates of the impact of adequate care on birth weight using data from the birth certificate (117 g) or medical records (68.2 g) were significantly lower than the estimate produced for those White women who had agreement between the birth certificate and medical records files on their Kessner Index score (140 g). These findings suggested that estimates of the effect of adequate care on birth weight unadjusted for measurement error could be underestimating the effect by at least 23 g (or less than 1 ounce). The Kessner Index results for Black women show a striking difference in estimates based on the birth certificate and medical record, with the intermediate and adequate care variables based the birth certificate showing a much stronger association with birth weight. Given the weak results from the medical record data, it is not surprising that the estimates based on the matching scores show little improvement over those based on the birth certificate data alone. In fact, the estimated impact of adequate prenatal care utilization on birth weight for Black women with matching Kessner scores across the two data sources (140 g) actually was less than the estimate produced from birth certificate data (188 g). The Sample 2 estimates for White women (results not shown), based on cases in which the Kessner Index score was the same across all three

Table V. Grams of Birth Weight Associated with Each Month of Delay in Prenatal Care Initiation by Race

Sample 1	OLS BC	OLS MR	TSLS MR-BC
Grams of birth weight per month of delay			
Whites	−10.2*	−0.295	−17.2*,+
Blacks	−19.9*	6.60	−36.8*,+

Sample 2	OLS BC	OLS MR	OLS MQ	TSLS MR-BC	TSLS MQ-BC
Grams of birth weight per month of delay					
Whites	−9.38	−3.02	−6.79	−16.7	−28.7
Blacks	13.3	4.24	6.77	24.6	57.5

Note: The label BC refers to the birth certificate report of prenatal care visits, while MR and MQ refer to the medical record and mothers' questionnaire reports, respectively. In the TSLS estimation, the heading MR-BC refers to the MR report instrumented with the BC report. The regression analysis also included gestational age, three indicators for mother's level of education, three indicators for father's education, mother's age and its square, total birth order, an indicator for a previous termination, and metropolitan residence.
*$p < .05$; **$p < .01$; $^{+}p < .05$ for Durbin–Wu–Hausman test, comparing TSLS result from OLS result from birth certificate.

data sources, indicated that moving from inadequate care to intermediate care yields an increase of 151 g (or 5.3 ounces) of birth weight, and moving from inadequate to adequate care yields an increase of 243 g (or 8.5 ounces) in birth weight. These estimates were 35% higher for adequate care and 37% higher for intermediate care than results not adjusted for measurement error.

The results from analyses using the APNCU Index as the measure of prenatal care (not shown) are similar to those from the Kessner Index, including the finding that the birth certificate report was the most strongly related to birth weight of the three sources. In addition, there was some evidence of measurement error attenuation bias in the APNCU Index; that is, the estimated beneficial effects of prenatal care were largest in those cases where the APNCU classification matched across the data sources. For example, for White women, the effect of adequate care on birth weight was estimated to be 74.4 g for those who APNCU scores are the same in the birth certificate and medical record files, compared with 55.2 g from the birth certificate data alone and 19.7 g from the medical record data alone. For Black

Table VI. Grams of Birth Weight Associated with Degree of Matching of Kessner Index Categories Between Birth Certificate and Medical Records by Race

	Whites			Blacks		
Kessner Index	BC	MR	BC-MR	BC	MR	BC-MR
Inadequate	—	—	—	—	—	—
Intermediate	48.5	9.55	—	161**	31.0	—
Adequate	117**	68.2**	—	188**	−14.8	—
Birth certificate category/Medical records category:						
Inadequate/inadequate	—	—	—	—	—	—
Inadequate/intermediate	—	—	−18.7	—	—	26.0
Inadequate/adequate	—	—	77.0	—	—	−104.0
Intermediate/inadequate	—	—	54.5	—	—	168.0
Intermediate/intermediate	—	—	**45.4**	—	—	**161.0****
Intermediate/adequate	—	—	84.1*	—	—	118.0
Adequate/inadequate	—	—	122.0*	—	—	294.0**
Adequate/intermediate	—	—	95.8**	—	—	195.0**
Adequate/adequate	—	—	**140.0****	—	—	**140.0***

Note: BC, birth certificate; MR, medical record.
*$p < .05$; **$p < .01$.

women, the effects of adequate care on birth weight were estimated to be 210 g for those whose APNCU scores are the same in the birth certificate and medical record files, compared with 156 g from the birth certificate data alone and 85.6 g from the medical record data alone.

An obvious difference between the results for the two prenatal care use indices is the presence of the additional category of "adequate plus" on the APNCU index, which has a negative association with birth weight. This group is comprised of women who initiated care early with more than the recommended number of visits for a pregnancy of their gestation. Thus, the negative effect of adequate-plus care was likely the result of the pregnancy complications and/or maternal medical conditions that caused these mothers to receive additional care. For both White and Black women, the negative impact of adequate-plus care on birth weight is greater among women with matching scores across data files than it is using just birth certificate or medical records data. For example, adequate-plus care resulted in a 143-g decrease in birth weight for White women (Sample 1) with matching APNCU scores compared with a 119-g decrease using birth certificate data and a 47.8-g decrease using medical records data.

Overall, the results of our numerous analyses suggested that prenatal care utilization was measured with a great deal of error in the NNS, and that, as a result, estimates of the effect of prenatal care on birth weight were downwardly biased or attenuated. The degree of attenuation bias varied across measures of prenatal care utilization, and the results varied somewhat for White versus Black mothers. In many cases, correcting for attenuation bias significantly increased the estimated effect of the prenatal care measure on birth weight. However, even our estimates that correct for attenuation bias, which were often double those from OLS, were not of a magnitude that clearly suggested that prenatal care had a major impact on birth weight. For example, our results adjusted for measurement error suggested that each month of delay in the initiation of prenatal care contributed to a 17.2-g decrease in birth weight for Whites and a 36.8-g decrease in birth weight for Blacks. This means that if care was delayed a total of 4 months, this would lead on average to a 68.8-g (2.4-ounce) reduction in birth weight for White infants and a 147-g (5.2-ounce) reduction in birth weight for Black infants. Similarly, our adjusted results suggested that moving from inadequate to adequate care on the APNCU Index was associated with a 74.4-g (2.6-ounce) increase in birth weight for White infants and a 210-g (7.3-ounce) increase for Black infants. Whether or not gains of this magnitude are clinically significant remains unclear.

DISCUSSION

In summary, the results presented in this paper indicate that prenatal care utilization is hard to measure. First, we found a considerable degree of disagreement across three important sources of information on prenatal care utilization. Second, we found that the discrepancies in the reports were related to some of the characteristics of the mother—Black race, metropolitan county residence, being unmarried—that also are related to low birth weight. Third, in OLS birth weight equations, we found—for virtually every measure of prenatal care utilization—that the birth certificate report of prenatal care use was a stronger predictor of birth weight than either the medical record or maternal questionnaire. These findings are consistent with the interpretation that attenuation bias, and hence error in prenatal care measurement, is least severe for information from the birth certificate report (in the case of the NNS data). Finally, we found evidence of attenuation bias from measurement error in estimates of the impact of prenatal care use on birth weight for both White and Black mothers. These results suggest that estimates of the efficacy of prenatal care utilization that are not corrected for measurement error will be biased downward (i.e., they will underestimate the beneficial impact of prenatal care). As mentioned above, however, the clinical significance of this underestimation (which is generally in the magnitude of a few ounces) is not clear.

Our results also suggest that, contrary to what is reported in several previous studies, much of the difference between TSLS and OLS estimates of the effect of prenatal care may be the result of attenuation bias in the OLS estimates rather than the results of selection bias only. Since we did not have geographic identifiers, we could not match our data with local medical care availability and other community-level variables that could act as instruments for the demand for prenatal care. To the extent that there is adverse selection in the demand for prenatal care, our estimates of the efficacy of prenatal care utilization are too low. However, our results indicate that estimators that explicitly account for selection bias but which are not robust to measurement error will substantially underestimate the effect of prenatal care. For example, consistent estimation with the selection-correction methodology suggested by Joyce (32) depends on the absence of misclassification in the Kessner Index. Our results suggest that these results are substantially attenuated.

There are several limitations to our study. First, as with most research on prenatal care, our study focused on prenatal care *utilization*, neglecting other dimensions of prenatal care, including the content, quality, continuity, and comprehensiveness of care. The effect of prenatal care on birth outcomes may vary by its adequacy in regard to its content and quality in addition to the adequacy of the number and timing of visits. Second, we did not have a measure of prenatal care utilization that can be considered the "gold standard." As stated above, our OLS results are consistent with the interpretation that attenuation bias is least severe for prenatal care information from the birth certificate. However, more definitive statements about the relative accuracy of the different data sources are not possible without a gold standard to which we can appeal. Also, due to its administrative nature, we used the birth certificate as an instrument in our TSLS estimates under the assumption that the error in this information on prenatal care is unsystematically related to birth weight. Again, however, without information from a gold standard, we cannot test this assumption. Third, an unbiased estimate of the partial correlation between prenatal care utilization and birth weight depends on the noncorrelation of the measurement errors across sources of prenatal care data used in the estimation. Since the raw correlations of the measures of prenatal care across sources are quite low, there appears to be a good deal of independence in the errors. However, if there is a common element in the errors in both sources, which would induce a positive correlation in the error terms, our TSLS estimates will be attenuated themselves, and we will have understated the extent of the attenuation bias in the OLS estimates. Fourth, since geographic identifiers for individuals in the NNS were not available to us, we were not able to match the birth records to area-level data on medical care availability and other area-level characteristics, preventing us from obtaining a second set of TSLS estimates robust to both selection and attenuation bias. If this second set of TSLS estimates were available, we could compare them to the TSLS and OLS presented here, giving us an understanding of the relative importance of attenuation bias and selection bias that is present in OLS estimates.

Fifth, the part of this study that incorporated maternal self-report information (Sample 2) is limited by the fact that only married women received the maternal questionnaire, that there was a high level of imputed prenatal care data in this data source, and that there was a small number of Black women in the resulting study sample. Thus, the generalizability of the results from Sample 2 is limited. Sixth, prenatal care use is not the only variable used in birth outcomes research that may suffer from measurement error. In particular, gestational age also is difficult to measure (36, 37). We do not, however, believe that errors in reporting of gestational age at delivery are bigger than errors in prenatal care (Table I).

A final limitation of this study is that the data are 20 years old. Nonetheless, at the time we began this study, the NNS was the only nationally representative data source with three different measures of prenatal care use for sample respondents. Although there have been some changes in obstetrical technology in the past two decades, recommendations regarding the content and the timing of standard prenatal care are virtually the same (14). Thus, we believe that our major conclusions are valid and important for both researchers and policymakers.

CONCLUSIONS

The relationship between prenatal care use and birth weight is difficult to measure, and two of the most complicated methodological issues (measurement error and selection bias) may lead to underestimations of the effect of care when left unaddressed. The NNS data provide an opportunity to evaluate the quality of the administrative and self-reported data on which most prenatal care utilization research is based. Our results show that the extent of measurement error in administrative records and self-reports of prenatal care utilization is substantial. Thus, studies that do not address measurement error in utilization—nearly every study done to date—will almost certainly underestimate the effect of prenatal care. A number of studies have addressed selection bias, but not all of the proposed methods are robust to measurement error. Therefore, our work helps define which selection-bias approaches are profitable avenues for future research.

A number of research recommendations come out of this study. We recommend that future studies of the association between prenatal care utilization and birth weight attempt to control for measurement error by using TSLS analysis when multiple sources of utilization data are available. When these data are not available, authors should acknowledge the probable presence of attenuation bias in the range and magnitude described here. We also recommend that, whenever possible, multiple independent measures of key variables that are likely to be measured with error be incorporated in future national natality or infant health data sets. This recommendation further extends to other types of health care utilization data where self-reports and/or administrative data have not been validated against a gold standard or where such a standard does not exist.

We have shown that three standard methods of collecting prenatal care utilization data are prone to substantial measurement error. Our research uncovers several important problems in retrospective reporting of prenatal care utilization, problems that are likely to exist in other areas of outcomes research as well. Future research must address both the problems of measurement error and selection bias if we are further to elucidate the relationship between prenatal care utilization and birth outcomes. An improved understanding of the potential of prenatal care to reduce the risk of low birth weight, prematurity, and infant mortality is essential to the further development of public health policies aimed at improving population reproductive outcomes.

While prior studies may have underestimated the potential impact of prenatal care use on birth weight outcomes, it is important to emphasize that our results (which take into account attenuation bias due to measurement error) do not offer unequivocal support for a continued focus on standard prenatal care as a major public policy response to the serious infant health problems in the United States. Although methodological problems preclude a precise estimation of the impact of prenatal care use on birth weight and other pregnancy outcomes, a growing body of evidence suggests that standard prenatal care is not a major determinant of infant health at birth. Such evidence led Huntington and Connell (8) to state that the strong and continued focus of public health policies on prenatal care implies "that there is a simple medical remedy for problems that are probably manifestations of deeply rooted social and economic factors." To be sure, better data and improved measurement approaches would lead to a better understanding of the true relationship between prenatal care and birth outcomes. However, the evidence to date does not suggest that, even if measured perfectly, standard prenatal care should emerge as the cornerstone of maternal and infant policy in the United States.

REFERENCES

1. Singh GK, Yu SM. Infant mortality in the United States: Trends, differentials, and projections, 1950–2010. *Am J Public Health* 1995;85:957–64.
2. Paneth N.S. The problem of low birth weight. *Future Child* 1995;5(1):19–34.
3. Sardell A. Child health policy in the U.S.: The paradox of consensus. *J Health Politics Policy Law* 1990;15:271–304.
4. Kotch JB, Blakely CH, Brown SS, Wong FY, editors. *A pound of prevention: The case for university maternity care in the U.S.* Washington, DC: American Public Health Association, 1992.
5. Alexander GR, Korenbrot CC. The role of prenatal care in preventing low birth weight. *Future Child* 1995;5(1):103–20.
6. Schlesinger M, Kronebusch K. The failure of prenatal care policy for the poor. *Health Affairs* 1990;4:91–111.
7. Fiscella K. Does prenatal care improve birth outcomes? A critical review. *Obstet Gynecol* 1995;85(3):236–44.
8. Huntington J, Connell FA. For every dollar spent: The cost savings argument for prenatal care. *New Engl J Med* 1994; 331(19):1303–07.
9. Kramer MS. Determinants of low birth weight: Methodological assessment and meta-analysis. *Bull World Health Org* 1987;5:663–737.
10. Kallan JE. Race, intervening variables, and two components of low birth weight. *Demography* 1993;30:489–506.
11. Cramer JC. Racial and ethnic differences in birth weight: The role of income and financial assistance. *Demography* 1995; 32:231–47.
12. Collins JW, David RJ. Differences in neonatal mortality by race, income and prenatal care. *Ethnicity Disease* 1992; 2:18–26.
13. Emanuel I, Hale CB, Berg CJ, Poor birth outcomes of American Black women: An alternative explanation. *J Public Health Policy* 1989;Autumn:299–306.
14. Institute of Medicine. *Preventing low birthweight.* Washington, DC: National Academy Press, 1985.
15. Alexander GR, Cornely DA. Prenatal care utilization: Its measurement and relationship to pregnancy outcome. *Am J Prevent Med* 1987;3:243–53.
16. Peoples-Shep MD, Kalsbeek WD, Siegel E. Why we know so little about prenatal care nationwide: An assessment of required methodology. *Health Services Res* 1988;23:359–80.
17. Frick KD, Lantz PM. Selection bias in prenatal care utilization: Linking economic and health services research. *Med Care Res Rev* 1996;53(4):371–96.
18. Alexander GR, Tompkins ME, Peterson DJ, Weiss J. Source of bias in prenatal care utilization indices: Implications for evaluating the Medicaid expansions. *Am J Public Health* 1991;81:1013–16.
19. Buescher PA, Taylor KP, David MH, Bowling JM. The quality of the new birth certificate data: A validation study in North Carolina. *Am J Public Health* 1993;83:1163–65.
20. Piper JM, Mitchel EF, Snowden M, Hall C, Adams M, Taylor P. Validation of 1989 Tennessee birth certificates using maternal and newborn hospital records. *Am J Epidemiol* 1993; 137:758–68.
21. Clark K, Fu CM, Burnett C. Accuracy of birth certificate data regarding the amount, timing, and adequacy of prenatal care using prenatal clinic medical records as referents. *Am J Epidemiol* 1997;145:68–71.
22. Forrest JD, Singh S. Timing of prenatal care in the United

States: How accurate are our measurements? *Health Services Res* 1987;22:235–53.
23. Fingerhut LS, Kleinman JC. Comparability of reporting between the birth certificate and the 1980 National Natality Survey. *Vital Health Stat* 1985; series 2(99):1–33.
24. Kogan MD, Martin JA, Alexander GR, Kotelchuck M, Ventura SJ, Frigoletto FD. The changing pattern of prenatal care utilization in the United States, 1981–1995, using different prenatal care indices. *JAMA* 1998;279;1623–28.
25. Frick KD, and Lantz PM, How well do we understand the relationship between prenatal care and birth weight? *Health Services Res* 1999;34(5):1063–73.
26. Keppel KG, Heuser RL, Placek PJ, *et al.* Methods and response characteristics, 1980 National Natality and Fetal Mortality Surveys. *Vital Health Stat* 1986;2(100):1–17.
27. Kessner DM, Singer J, Kalk CE, Schlesinger ER. *Infant death: An analysis by maternal risk and health care.* Washington, DC: Institute of Medicine and National Academy of Sciences, 1973:Chapter 2.
28. Gortmaker SL. The effects of prenatal care upon the health of newborn. *Am J Public Health* 1979;69(7):653–59.
29. Kotelchuck M. An evaluation of the Kessner adequacy of prenatal care index and a proposed adequacy of prenatal care utilization index. *Am J Public Health* 1994;84:1414–20.
30. Kotelchuck M. The adequacy of prenatal care utilization index: Its US distribution and association with low birthweight. *Am J Public Health* 1994;84:1486–89.
31. Ashenfelter O, Krueger A. Estimates of the economic return to schooling from a new sample of twins. *Am Econ Rev* 1994;84:1157–73.
32. Joyce TJ. Self-selection, prenatal care, and birthweight among Blacks, Whites, and Hispanics in New York City. *J Hum Resources* 1994;29(3):762–94.
33. Card D. The effect of unions on the structure of wages: A longitudinal analysis. *Econometrica* 1996;64(4):957–79.
34. Davidson R, MacKinnon JG. *Estimation and inference in econometrics.* New York: Oxford University Press, 1993: 237–42.
35. Rosenzweig MR, Schultz TP. The stability of household production technology: A replication. *J Hum Resources* 1988; 23:535–49.
36. Lantz PM, Partin M. Population indicators of prenatal and infant health. In: Hauser RM, Brown BV, Prosser WR, editors. *Indicators of children's well-being.* New York: Russell Sage Foundation, 1997.
37. David RJ. The quality and completeness of birth weight and gestational age data in computerized birth files. *Am J Public Health* 1980;70:964–73.