Probability Samples of Area Births Versus Clinic Populations for Reproductive Epidemiology Studies

David A. Savitz,[1,2] Nancy Dole,[1] Diane Kaczor,[1] Amy H. Herring,[3] Anna Maria Siega-Riz,[1,4] Jay Kaufman,[1,2] John M. Thorp Jr[1,5]

[1]Carolina Population Center, University of North Carolina, Chapel Hill, North Carolina

[2]Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, North Carolina

[3]Department of Biostatistics, University of North Carolina School of Public Health, Chapel Hill, North Carolina

[4]Departments of Maternal and Child Health and Nutrition, University of North Carolina School of Public Health, Chapel Hill, North Carolina

[5]Department of Obstetrics and Gynecology, University of North Carolina School of Medicine, Chapel Hill, North Carolina

Reprint requests and correspondence to:

Professor David A. Savitz
Department of Epidemiology, CB #7435
University of North Carolina School of Public Health
Chapel Hill, NC 27599-7435
Telephone: 919-966-7427
Fax: 919-966-2089
E-mail: david_savitz@unc.edu

SUMMARY

Studies of pregnancy outcome are generally either based on geographically defined populations, often from birth records, or on clinic or hospital populations. We compared women recruited into the Pregnancy, Infection, and Nutrition (PIN) Study with women who resided in the geographic area of the study (Alamance, Orange, and Wake Counties in North Carolina) and gave birth over the corresponding time period (1996-2000). Clinic participants were more likely to be Black, younger, have lower education, be unmarried, have a more frequent history of adverse pregnancy outcome, obtain prenatal care later, and smoke more cigarettes. Despite that profile, the proportion of PIN participants delivering preterm was somewhat lower than among area women overall (10.8% vs. 11.3%). Black/white risk ratios for preterm birth were markedly different for area (1.6) versus PIN women (1.1), whereas other predictors were similar. Patterns may differ across groups for many reasons, including self-selection of clinics and varying clinical practices.

INTRODUCTION

Research on determinants of pregnancy outcome generally relies on one of two population sources: either geographic rosters based on maternal residence or selected health care providers, typically prenatal clinics or hospitals of delivery. Geographically defined populations are often studied using birth records,[1,2] which provide large numbers and a comprehensive roster of births (or the opportunity to sample from such a roster). Alternatively, such populations can be surveyed systematically as is done in the US by the National Center for Health Statistics[3] or by recruiting a cohort from a geographically defined population for follow up.[4] Such geographically defined populations are of direct demographic and public health interest in that they constitute the relevant level of analysis for questions of public policy applied to politically defined units such as states or counties.

Such studies are rarely able to obtain detailed information on determinants of health and refined outcomes beyond what is included in the birth certificate or in other public records. More detailed information comes only from intensive interviews,[5] collection of biological specimens,[6,7] or from special clinical procedures such as ultrasound examinations.[8] Rarely, and with great effort, probability samples can be chosen from birth records in order to obtain more detailed information that can be directly extrapolated to the total population, e.g., the National Maternal and Infant Health Survey.[9] However, even then the amount of information that can be obtained is often limited to interviews or medical record reviews and substantial non-response results in the potential for bias. Obtaining needed access to individuals and their medical care providers is a daunting, perhaps insurmountable challenge for geographically defined populations.

In contrast, studies in clinical settings such as prenatal care practices allow for much more intensive data collection efforts.[5-7,10] Advantages include the ability to identify participants early in their pregnancy and monitor the pregnancy through its course and through childhood, collection of biological specimens, and extensive and repeated interviews as desired with the patient and/or health care provider. The aggregation of pregnancies and births in specific health settings (prenatal clinics, hospitals) facilitates this type of research. The major loss in such research is representativeness of the population residing in the geographic area. Even though epidemiologists speak of the "population served by the hospital or clinic," it is only the receipt of services at the hospital or clinic that defines the collection of individuals. Obtaining such services in the US health care system, which is not organized to provide services for geographically defined populations, isolates a nonrandom subset of the otherwise eligible population residing in the general area that is unlikely to be representative in regard to demographic characteristics and health outcomes.

In order to make more informed choices regarding the strengths and limitations in these competing research strategies for programs such as the National Children's' Study as well as for more circumscribed investigations, empirical data on the inferences that can be drawn from clinical and geographically defined populations would be useful. Although clinical populations can be made more representative through targeted selection of specific clinics or hospitals, and probability samples of the population can be enhanced with more intensive data collection, there will always be a significant gap between these two strategies. This issue has been a source of controversy in planning for the National Children's' Study – the tradeoffs in what types of data can be collected and how broadly the population should be constituted.[11] To help inform such decisions, we have examined influences on pregnancy outcome in a clinic population enrolled as part of the Pregnancy, Infection, and Nutrition (PIN) Study and compared results for our clinical population to results for births in the counties served by the participating clinics. We examined the characteristics of the women served by those clinics, their pregnancy outcomes, and the degree to which relationships between predictors and outcomes are consistent between clinical and area populations.

METHODS

The PIN Study was conducted in central North Carolina,[10] enrolling women at the Wake County Human Services/Wake Area Health Education Center prenatal clinics in Raleigh and at the major obstetrics clinics at the University of North Carolina Hospital in Chapel Hill. These providers, based in Wake and Orange counties of North Carolina, respectively, serve a population that comes largely from Orange, Wake, and Alamance counties. We recognize that within the counties served, geographic proximity to the participating clinics may affect likelihood of selecting those sites for prenatal care, but counties were the units most feasible for the analyses of interest.

Births to women enrolled in the PIN Study during the period 1996-2000 and all live births from Alamance, Orange, and Wake counties over the same time period were evaluated. In addition to these extreme comparisons of PIN participants versus all area women who gave birth, we considered two intermediate groups. To isolate the effect of non-response from clinic selection, we examined PIN-eligible women, regardless of whether they did or did not participate in the study. We had sufficient information on those women from medical charts to evaluate most predictors of interest. On the other end, we considered a restricted area sample in which we imposed some of the requirements for being enrolled in the PIN Study, specifically age 16 or over, prenatal care onset by the seventh month of pregnancy, and non-Hispanic ethnicity (since PIN protocols require the ability to be interviewed in English). Thus we had four groups to consider, ordered by increasing restrictiveness as area women, restricted area women, PIN eligible women, and PIN participants. However, without exception, PIN participants and PIN-eligible women yielded indistinguishable results, as did area women with or without restrictions (data not shown). We chose to present detailed results comparing PIN participants and restricted area women as the most informative comparison for addressing the methodologic issues in these types of studies.

The primary pregnancy outcome of interest was preterm birth (<37 weeks' completed gestation), the focus of the PIN Study. Gestational age in the PIN Study was based on an algorithm that relied on last menstrual period unless ultrasound prior to 20 weeks' gestation resulted in a deviation of 14 days or more in estimated date of conception, in which case the ultrasound estimate was substituted. In vital records, dates are based on reported last menstrual period, unless missing, in which case the physician's estimate is used. Subject to the limited numbers of events, we also examined more severe (<34 weeks) preterm birth as a separate outcome and term low birthweight (term-LBW), a measure of reduced fetal (defined as gestational age of 37 completed weeks or longer and birthweight less than 2500 grams).

Predictors of pregnancy outcome were race (White, Black, other), mother's age (<20, 20-29, 30-34, 35+), education (<12, 12, 13-15, 16+ years), marital status (married, not married), pregnancy history (no prior births, 1+ prior births/none preterm or small-for-gestational-age, 1+ prior births/history of preterm or small-for-gestational-age), month of entry into prenatal care (<3, 3-4, 5+), prior preterm or small-for-gestational-age birth (no, yes), and smoking (none, 1-9, 10-19, 20+ cigarettes per day). Only those predictors available for both area women and PIN women could be considered, limiting the scope of the analysis.

First, we examined the distribution of the predictor variables, comparing area and PIN populations. Next, we considered the risk of the pregnancy outcomes among subsets of women in each of the populations, e.g., the proportion of area Black women who delivered preterm compared to the proportion of PIN Black women who delivered preterm. Then, we characterized the association between predictors and outcomes in each of the groups, using the risk ratio, e.g., the ratio of the risk of preterm births among area Black women compared to area White women, and the ratio of the risk of preterm birth in PIN Black women compared to PIN White women. This was first calculated without adjustment for other factors and then, using log-linear modeling, with adjustment for all the other predictors listed above. For the less frequent outcomes (preterm birth <34 weeks, term-LBW), multivariate analysis was not feasible.

In order to evaluate the observed differences for predictors of interest between the PIN participants and the general population, we fit a combined model allowing separate effects of each factor for PIN women and women in the restricted 3 county area and contrasted it, for each predictor, to a model in which the coefficients were constrained to be the same. For example, we fit two separate indicators of "not married" status -- one for PIN women only, and another for women in the restricted area population only -- and generated a p-value for the hypothesis that marital status had the same effect on the risk of preterm birth for women in the PIN and in the general population.

In the study area over the time period of interest, there were 59,979 births in total, though for the analyses of preterm birth, the numbers were reduced modestly. The imposition of the eligibility restrictions used by the PIN Study reduced the area population by 9% to 53,436. PIN eligible women totaled 3,820, or 6.4% of the total area population (6.5% of Alamance County births, 15.0% of Orange County births, and 5.2% of Wake County births). Of those eligible, 60% were recruited, for a total of 2,289 women.

RESULTS

The social and demographic distribution of births in the area and to PIN participants differed markedly (Table 1). Relative to area births, PIN participants were far more likely to be Black (43% vs. 23%), younger, have lower education, be unmarried, have a more frequent history of adverse pregnancy outcome, obtain prenatal care later, and smoke more cigarettes. Across all these factors, PIN participants had a less favorable risk profile relative to area women. Despite that profile, however, the proportion of PIN participants delivering preterm was somewhat lower than among area women overall (10.8% vs. 11.3%) and the proportion that were term-LBW very similar (2.8% vs. 2.6%). Such a pattern is suggestive of differing risk of adverse outcomes *within* risk factors subsets.

With regard to race and preterm birth (Table 1), White PIN participants had slightly greater risk than their area counterparts, whereas Black PIN participants had notably lower risk of preterm birth than their area counterparts, 11.1% versus 16.7%. Differences in term-LBW were far more modest but in the same direction. It should be noted that the unfavorable risk factor profile described above for the total PIN population applied similarly to White and Black subsets (data not shown), so that the favorable outcomes in Black PIN participants is not a function of a their having a more favorable risk factor profile.

The advantage of PIN women in preterm birth is concentrated among the youngest mothers, with some increased risk among older women in the PIN Study (Table 1). Similarly, the reduced preterm birth risk among PIN women is exclusively among women with 12 or fewer years of education, those not married, nulliparous women, women with later onset of prenatal care, and non-smokers or light smokers. In most cases, the same general pattern occurs for term-LBW, though the relative magnitude for preterm birth and term-LBW differs across the predictors. The aggregate differences in pregnancy outcome found for area versus PIN women are not simply a function of differing risk factor profiles, but rather due to a complex pattern of differing outcome risks *within* subgroups combined with differing population composition, particularly among the subgroups expected to have higher risk.

The finding that these particular clinics serve a distinct, non-random subset of the women residing in the area is not surprising, and perhaps it is not too surprising that outcomes differ within ostensibly homogeneous subsets of women. A critical question is whether the pattern of associations with pregnancy outcomes differs across the two data sources. That is, does patient source *modify* the association between predictors and outcome? With regard to race, effect-modification is clearly present (Table 2). The adjusted Black/White risk ratio is 1.6 for preterm birth for area women versus 1.1 for PIN participants (p=0.04), with the difference even more pronounced for births <34 weeks' gestation (RR=2.3 [95% CI = 2.0, 2.5] for area women versus RR=1.0 [95% CI = 0.5, 2.1] for PIN participants). In contrast, the Black/White risk ratio for term low birthweight is somewhat greater for PIN participants (RR = 2.4 [95% CI = 1.1, 5.4]) than for area women (RR = 1.8 [95% CI = 1.5, 2.0]).

The pattern of the association of preterm birth with age, education, marital status, prior birth outcome, prenatal care onset, and tobacco use is similar for PIN and area women, with all p-values for the contrast equal to 0.44 or greater. In most cases (except for prior adverse birth outcome) the risk ratios are modestly different from 1.0 and there is a problem of imprecision in the estimates for PIN participants. In our attempt to evaluate more severe preterm birth (<34 weeks) and term low birthweight, the problems of precision were much more severe. Taking that uncertainty into account, no striking differences were found in the pattern of risk ratios among area and PIN women (data not shown), with the exception of the association between smoking 10+ cigarettes per day and more severe preterm birth where the risk ratio was 1.5 [95% CI = 1.3, 1.8] for area women and 0.7 [95% CI = 0.2, 2.3] for PIN participants.

DISCUSSION

The pattern of preterm delivery among women enrolled from selected prenatal clinics into the PIN Study differed from women giving birth in the area in several ways. The social, demographic, and behavioral profile of the specific clinics were not reflective of the composition of the area population, not surprising given the many factors determining clinic attendance. Basing the study in a county health department and university medical center resulted in a clear shift towards socially disadvantaged women, as reflected in the higher proportion who were Black, of lower education, unmarried, had later prenatal care onset, a greater frequency of adverse prior pregnancy outcome, and smoked during pregnancy. If the lack of representativeness on such readily measurable traits were the only difference between the clinic and area population, the PIN participants could simply be viewed as a weighted or stratified sample from the area (e.g., with a higher sampling weight for Blacks than Whites), and the results from the PIN Study could be statistically extrapolated for application to the area population. However, the pattern is far more complex in that the frequency of preterm delivery appears to differ between clinic and area women within strata of risk predictors.

The similar aggregate risks of adverse outcomes for PIN participants and eligible area women despite an unfavorable risk factor profile among the study participants is indicative of lower risks of adverse outcome within strata. In particular, a consistent pattern was found in which the PIN participants in the highest risk subgroup (Blacks, lower education, unmarried, later prenatal care onset, prior adverse pregnancy outcome) had lower risk than their area counterparts, with similar risks of preterm birth in the lower risk subgroups. It would seem that the selectivity for attending these clinics and being eligible to enroll in a research protocol is much greater among disadvantaged women. There are several possible reasons such a pattern might be found. Among women in certain strata, e.g., young, unmarried, perhaps only some are sufficiently motivated to seek the health services available through the university or county health department, reflective of other (unmeasured) favorable predictors of pregnancy outcome. Alternatively, perhaps the intensive, high quality prenatal care provided at these sites is truly beneficial, but only among those who are disadvantaged and in real need of the medical and social benefits of such care. Women of greater means may do well regardless of the source of care. Another possibility is that more advantaged women in the area who are not attending the study clinics are obtaining similar quality of care elsewhere, whereas less advantaged women not attending these clinics are obtaining less effective prenatal care services elsewhere. Complicating the issue of addressing clinical services is the iatrogenic nature of many preterm births, in which early delivery is the result of clinician intervention, with the goal of avoiding more undesirable outcomes such as eclampsia or stillbirth. Within the PIN Study, 43% of preterm births were medically indicated, a far greater proportion of those at the university medical center (76%) than those at the county health department (25%), whereas the distribution of spontaneous and medically indicated preterm births among area women is not known. Observing broadly similar patterns for term-LBW provides some reassurance that the entire picture is not a product of differing clinician decision-making regarding early delivery in the study settings.

With regard to the patterns of association within the PIN Study and the area population, the results are mixed. The most striking difference is seen for Black/White differences, where the Black and White women who are in the study are much more similar to one another in their risk of preterm birth than Black and White women in the area, largely due to the anomalously favorable pregnancy outcomes among Black women. Within the limitations of the data (noted below), the patterns of association between other predictors and pregnancy outcome comparing the two populations were quite similar. This provides limited assurance that inferences from the PIN Study and similarly clinic-based studies may be applicable more broadly, and that patterns found in broader populations may apply in particular subgroups defined by clinic attendance. However, this is a tenuous inference given how weak the predictors are overall and the notable exception found for Black/White comparisons.

A number of limitations regarding both internal and external validity of these findings should be noted. First, with regard to internal validity, there are differences in the methods and quality of assignment of gestational age in comparing the PIN and area populations. In the PIN Study, we used last menstrual period information in combination with early ultrasound, assigning dates based on last menstrual period unless the deviation from ultrasound dates was 14 days or greater, in which case we relied on the ultrasound. This had a marked impact on postterm delivery,[8] but little impact on the frequency of preterm birth. Nonetheless, this difference may well have influenced the pattern of preterm birth, and to a lesser extent, the pattern of term-LBW. Second, the imprecision in the PIN Study population analyses is a serious concern that could readily mask or create the false impression of associations. For relatively rare exposures, e.g., history of prior preterm or SGA birth, smoking 20+ cigarettes per day, the findings are quite imprecise. Third, there are few strong predictors of preterm birth, making inferences from the contrast of PIN and area populations problematic since all relative risks are so close to the null. The questionable quality of some birth certificate information relative to the data collected specifically for the PIN study further cloud the comparisons. For example, although the mother's smoking status is ascertained, there are reasons to question its quality[12] relative to the systematic assessment conducted for the PIN study. Even such items as race or education may differ from the two sources. Finally, the opportunity to delve more deeply into the reasons for the observed patterns is limited by the data available for area women on a wide range of potential risk factors (e.g., nutrition, stress, physical activity) and such important considerations as spontaneous versus medically indicated preterm birth, the details of prenatal care, the reasons for attending the study clinics, etc.

Obviously, if the goal of research were solely to characterize broad patterns in the population, relating social and demographic characteristics to pregnancy outcome and ultimately to children's health, the large sample size and scope of the geographically defined populations would make that approach superior. From that perspective, the patterns found for readily measured attributes such as education or maternal age are more valid for the area population (subject to the quality of outcome assessment). However, what this analysis does not reflect is the type of information that can feasibly and reliably be obtained in one approach versus another. The clinical settings in which the PIN study was conducted were chosen not for reasons of representativeness but because they serve selectively high-risk patients and were willing and able to accommodate a complex and demanding research protocol. What is not included in these analyses are the many potentially important factors that are available only through such effort, e.g., detailed assessment of psychosocial stress,[13] physical activity,[14] diet,[15] or cocaine use.[16] For studying biological mechanisms that require collection of specimens during the course of pregnancy, conducting ultrasound examinations, etc., there is no alternative to the selectivity that was imposed for the PIN study. Even for data that are ostensibly comparable, such as cigarette smoking, the quality of information attainable through detailed interviews is far superior to that available from vital records. For example, we were able to go beyond self-reported smoking at a given point in time and examine the changes throughout the course of pregnancy and collect urine at two times in order to assay cotinine, a metabolite of tobacco smoke that provides more definitive information regarding recent smoking.[17]

Thus, the different approaches to conducting studies of pregnancy and subsequent health of children and mothers have competing goals and compensating strengths and weaknesses. The existing vital records resources allow for monitoring general trends in reproductive health. The problem is that to refine and extend our understanding of those patterns calls for research that can delve more deeply into such issues of why risks for adverse birth outcomes are higher among Blacks or lower among Mexican-Americans compared to Whites. The competing strategies are to tackle such questions directly through probability sampling from the US population versus studying convenience samples intensively and then attempt to extrapolate back to the population. There are stark differences in these approaches, with clear strengths and weaknesses inherent in each. For specific goals, the optimal strategy becomes clear, but when there is a lengthy and heterogeneous list of goals as exist for the National Children's Study, the ultimate decision will ensure that some questions can be answered well and others poorly or not at all.

ACKNOWLEDGEMENTS

This study was supported by grant number HD28684 from the National Institutes of Health; and cooperative agreements S455-16/17, S0807-18/18, S1099-19/19 through the Association of Schools of Public Health/Centers for Disease Control and Prevention, U64/CCU412273 through the Centers for Disease Control and Prevention; and funds from the Wake Area Health Education Center in Raleigh, North Carolina.

REFERENCES

1. MacDorman MF, Mathews TJ, Martin JA, Malloy MH. Trends and characteristics of induced labour in the United States, 1989-98. *Paediatric and Perinatal Epidemiology* 2002;16:273-273.

2. Leonard H, Petterson B, Bower C, Sanders R. Prevalence of intellectual disability in Western Australia. *Paediatric and Perinatal Epidemiology* 2003;17:58-67.

3. Simpson GA, Colpe L, Greenspan S. Measuring functional developmental delay in infants and young children: prevalence rates from the NHIS-D. *Paediatric and Perinatal Epidemiology* 2003;17:68-80.

4. Decouflé P, Autry A. Increased mortality in children and adolescents with developmental disabilities. *Paediatric and Perinatal Epidemiology* 2002; 16:375-382.

5. Giannelli M, Doyle P, Roman E, Pelerin M, Hermon C. The effect of caffeine and nausea on the risk of miscarriage. *Paediatric and Perinatal Epidemiology* 2003;17:316-323.

6. Xiao R, Sorensen TK, Frederick IO, El-Bastawissi A, King IB, Leisenring WM, Williams MA. Maternal second-trimester serum ferritin concentration and subsequent risk of preterm delivery. *Paediatric and Perinatal Epidemiology* 2002;16:297-304.

7. Skalkidou A, Petridou E, Papathoma E, Salvanos H, Trichopoulos D. Growth velocity during the first postnatal week of life is linked to a spurt of IGF-1 effect. *Paediatric and Perinatal Epidemiology* 2003;17:281-286.

8. Savitz DA, Terry JW Jr, Dole N, Thorp JM Jr, Siega-Riz AM, Herring AH. Comparison of pregnancy dating by last menstrual period, ultrasound, and their combination. *American Journal of Obstetrics and Gynecology* 2002;187:1660-1666.

9. Sanderson M, Gonzalez JF. 1988 National Maternal and Infant Health Survey: methods and response characteristics. *Vital & Health Statistics* - Series 2: Data evaluation & methods research 1998; 125: 1-39.

10. Savitz DA, Dole N, Williams J, Thorp JM, McDonald T, Carter AC, Eucker B. Determinants of participation in an epidemiologic study of preterm delivery. *Paediatric and Perinatal Epidemiology* 1999;13:114-125.

11. Kaiser J. Everything you wanted to know about children, for $2.7 billion. 2003;301:162-163.

12. Webb DA, Boyd NR, Messina D, Windsor RA. The discrepancy between self-reported smoking status and urine cotinine levels among women enrolled in prenatal care at four publicly funded clinical sites. *Journal of Public Health Management and Practice* 2003;9:322-325.

13. Dole N, Savitz DA, Hertz-Picciotto I, Siega-Riz AM, McMahon MJ, Buekens P. Maternal stress and preterm birth. *American Journal of Epidemiology* 2003;157:14-24.

14. Evenson KR, Siega-Riz AM, Savitz DA, Leiferman JA, Thorp JM Jr. Vigorous leisure activity and pregnancy outcome: the Pregnancy, Infection, and Nutrition Study. *Epidemiology* 2002;13:653-659.

15. Siega-Riz AM, Bodnar LM, Savitz DA. What are pregnant women eating? Nutrient and food group differences by race. *American Journal of Obstetrics and Gynecology* 2002;186:480-486.

16. Savitz DA, Henderson L, Dole N, Herring A, Wilkins DG, Rollins D, Thorp JM Jr. Indictors of cocaine exposure and preterm birth. *Obstetrics and Gynecology* 2002;99:458-465.

17. Savitz, DA, Dole N, Terry, JW, Zhou H, Thorp JM. Smoking and pregnancy outcome among African-American and White women in Central North Carolina. *Epidemiology* 2001;12:636-642.