

National Breast and Cervical Cancer Early Detection Program: Methods to Determine Participation of Eligible Populations

Florence Tangka,* James Gardner,*
Brett O'Hara,** Joanna Turner,** and Mark Bauder**

* Centers for Disease Control and Prevention, ** U.S. Census Bureau

This report informs interested parties of ongoing research and to encourage discussion of work in progress. The views expressed in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention or the U.S. Census Bureau.

Background

Breast and cervical cancer screening rates among low-income, uninsured, and underinsured women are low (1, 2). Timely breast cancer screening with mammography and routine Papanicolaou (Pap) testing for cervical cancer have helped to reduce the incidence of these diseases as well as the mortality rate associated with them (3). Medical guidelines suggest that most women should have a screening mammogram every 1 to 2 years from ages 40 to 74 (4) and that they should begin cervical cancer screening within 3 years of the onset of sexual activity or age 21 (whichever is earlier) and have follow-up screening at least once every 3 years (5). Women ages 50-64 are at the highest risk for both breast and invasive cervical cancer, and 75 percent of all diagnosed cases of breast cancer are among women ages 50 years or older (6). Women without health insurance are less likely to follow screening guidelines for breast and cervical cancer and are at greater risk for a late diagnosis than are women with health insurance (7).

In response to the Breast and Cervical Cancer Mortality Prevention Act of 1990, the Centers for Disease Control and Prevention (CDC) established the National Breast and Cervical Cancer Early Detection Program (NBCCEDP) in 1991 as a means of ensuring access to and the quality of critical breast and cervical cancer screening services to underserved or uninsured women. The NBCCEDP provides funding, consultation and technical assistance to all 50 states and the District of Columbia (DC), as well as to 4 U.S. territories, and 13 American Indian/Alaska Native jurisdictions. The NBCCEDP gives grants to these entities to provide breast and cervical cancer screening services. Generally, those eligible for the NBCCEDP are women ages 18-64, who do not have health insurance coverage or whose insurance does not cover breast and cervical cancer screening and diagnostic services, and whose family incomes are at or below 250 percent of the federal poverty level. By law, the program services must be provided at no charge to women whose family income is below the federal poverty level. If these minimal services can be maintained, then individual programs are permitted to increase the income criteria. Nearly all programs set their income criteria at 200% or 250% of the FPL. Twenty-two states and DC currently set their income eligibility criteria at the lower

level. A high proportion of program participants represent women from minority populations and women isolated from existing services (8).

Because the NBCCEDP operates through a series of individual cooperative agreements with the grantees, the operational focus of individual NBCCEDP grantees may differ significantly. For instance, programs may direct their efforts based on area mortality trends or known high-risk sub-populations, and use different intervention strategies and organizational structures to carry out these efforts. The NBCCEDP grantees collect surveillance data on women served through the program and report data elements to the CDC that describe the demographic characteristics, screening history, and screening and diagnostic outcomes for these women. Since its inception, the program has screened over 2.3 million women and diagnosed 17,009 cases of breast cancer; 61,474 cases of precancerous cervical lesion; and 1,157 cases of cervical cancer (9).

Under the original authorizing legislation, the NBCCEDP could provide funds only for screening and diagnostic services, while individual programs had to secure other resources to treat the women with cancers diagnosed. The Breast and Cervical Cancer Prevention and Treatment Act of 2000 (amended in 2001) gave states the option to use Medicaid funding to provide medical assistance to eligible women who were screened by the NBCCEDP and found to need treatment.

The CDC is interested in estimating the number of women eligible for the NBCCEDP and the percentage of eligible women who are actually screened through the program. These estimates can be used for program evaluation, resource allocation, planning, partnership formation, outreach efforts, research, and the development of comprehensive cancer control initiatives. The Annual Social and Economic Supplement (ASEC) to the Current Population Survey (CPS) is widely cited for health insurance statistics and has a large sample. The CPS ASEC is used here to estimate the number of women eligible for the program. These direct estimates are provided both at the national and state levels. At the national level, CPS ASEC direct estimates are reliable for estimating the NBCCEDP eligible women by age, race, and ethnicity. However, the state-level CPS ASEC direct estimates are not reliable for low-income uninsured women by these characteristics.

National Estimates of the Percentage of NBCCEDP Eligible Women Who are Screened

The percentage of NBCCEDP-eligible women who are screened was calculated using the number of women who received NBCCEDP-funded mammograms and Pap tests during 2000-2003, combined with estimates of the number of women eligible for screening services from the CPS ASEC. Following the recommendations of the U.S. Preventive Services Task Force, the participation rates were estimated for women having a mammogram over a 2-year period (2002-2003) and for women having a Pap test over a 3-year period (2000-2002).

In 2002-2003, approximately 4 million (8.5%) of all U.S. women aged 40-64 met the NBCCEDP eligibility criteria. Of these women, 528,622 (13.2%) received an NBCCEDP-funded mammogram. During 2000-2002, about 10 million (11%) of all U.S. women aged 18-64 were eligible for the NBCCEDP, and 596,090 (6%) of them received at least one NBCCEDP-funded Pap test. Although the NBCCEDP provided breast and cervical cancer screenings to over a half-million low-income uninsured women, it served only a small percentage of the eligible population.

Although these national estimates of the number of women eligible for NBCCEDP-funded cancer screening and the percentage being screened can be helpful in planning outreach activities at the state level, reliable state and sub-state estimates of the eligible populations and the percentages of those populations actually screened is necessary in order to conduct operating-level program evaluations.

Estimating NBCCEDP Eligible Women at the County- and State-Level

The U.S. Census Bureau's project of estimating NBCCEDP eligible women builds on previous work on making model-based estimates of health insurance coverage. The Small Area Health Insurance Estimates (SAHIE) program was created to develop model-based estimates of health insurance coverage by age for all counties and states (10). Comprehensive and county-level data on health insurance coverage are not available from any federal survey. Even large national surveys such as the CPS ASEC do not have sufficient sample to provide direct estimates at the county level. The SAHIE program's focus is to produce a consistent set of nation-wide estimates for sub-state areas. The estimates are adjusted so that, before rounding, county numbers sum to their states and similarly the states sum to the CPS ASEC national estimates. This is a new program at the Census Bureau and the first ever set of estimates were released in July 2005.

To estimate detailed program participation rates, the NBCCEDP requires state- and county-level estimates of the number of uninsured women by race and ethnicity, age, and income. If these estimates are proven feasible and statistically reliable, the CDC can use them to evaluate the state-to-state and county-to-county variations in the demographic makeup of program participants. As noted previously, the definition of "low income" also varies by state, typically 200 or 250 percent of the poverty threshold. These categories correspond to the majority of women served by the individual NBCCEDP programs.

The number of women eligible for the NBCCEDP at the county-level cannot be calculated by direct survey estimates. At the state-level, direct estimates of low-income uninsured women by age group are available but have large variances. When the number of women is further grouped into income, age, race, and ethnicity categories, some categories at the state-level cannot be estimated by direct survey estimates. Direct estimates of health insurance coverage do not exist for most counties and the aforementioned problems are exacerbated.

The remainder of this paper focuses on the Census Bureau's evaluation of and methodology for estimating the number of eligible women for the NBCCEDP by state and county. A more detailed discussion can be found on the SAHIE program's web site (10). If these estimates reach production quality, they will be used by the NBCCEDP in program administration. The model-based estimates of health insurance coverage combine survey data with population estimates and administrative records and are based on data from the following sources: CPS ASEC; demographic population estimates; aggregated federal tax returns; aggregated food stamp records; and aggregated Medicaid participation records.

This project requires small area estimation procedures because of the small sample size for the demographic breakdowns and lack of data for many counties. A small area for which a quantity is estimated is often referred to as a domain. If the estimate of interest were the number of uninsured persons by state, there would be 51 domains. As the number of domains increase, estimation becomes more difficult. For the state estimates, domains are defined by age, race, sex, and Hispanic origin (ARSH) and IPR, and for the county estimates, domains are defined by age, sex and IPR. The categories that define the domains are the following: five age categories (0-17, 18-39, 40-49, 50-64, 65 and older), three race and Hispanic origin categories (Hispanic, non-Hispanic non-Black, non-Hispanic Black), two sex categories, and three IPR categories (0-200, 201-250, above 250). The estimates produced are for fewer domains per county because of smaller sample sizes in counties. There are 90 domains for each state and 30 domains for each county. The state model is used to produce estimates for a total of 4,590 domains and the county model for a total of 94,200.

For both the state and county model, Bayesian methods are used for the estimation. The Bayesian methods employed use Markov Chain Monte Carlo (MCMC) algorithms (11). Quantities which must be approximated in non-Bayesian methods, and for which the approximations are only known for certain sets of modeling assumptions, such as some standard errors (12), can be calculated using MCMC to the desired precision.

State-Level Model

In the state model the numbers of people (by ARSH) in the IPR categories and the health insurance rates within IPR categories are modeled jointly. The relationship of the IPR proportions to the ARSH characteristics is modeled with a multiple category logistic regression with normal errors; this is similar to the traditional multinomial regression. The distributions of proportions insured, conditioned on membership in the IPR categories, are modeled as logistic regressions with normal errors.

The fit of the models to the data is very important for the success of these models, and a substantial effort is made in checking the fit. Our primary means of assessing the validity of the models are: examining variances and coefficients of variation (CVs) of the estimates, posterior predictive p-values (ppp-values) (13), and other plausibility checks. We examine ppp-values constructed to evaluate the fit with respect to the means and variances, both for global behavior and for behavior within subgroups. Examination of these ppp-values does not show failures in the model with respect to the expectation of

the CPS ASEC or with respect to the variance. Variances and relative variances seem consistent with the conclusion that production of these estimates is feasible. For uninsured women between the ages of 40-64 and with incomes less than or equal to 200 percent of the federal poverty threshold, the CV is .23. While the results are promising, the fit of the current model can certainly be improved.

County-Level Model

The county model treats the numbers of people (by age and sex) in an IPR category and the proportions insured within the IPR category independently. The estimate of the number of people in each domain and the proportion insured within each domain are done separately, and the resulting estimates are then combined. The estimated proportion of uninsured (one minus the estimated proportion of insured) is multiplied by the estimated number of persons (from the IPR categories) to produce estimates of the number of persons uninsured for each domain.

Because IPR numbers and health insurance coverage proportions are estimated separately, two quantities are obtained to calculate total variance; a posterior variance for the number in the IPR category, and a variance for the proportion insured *conditioned on* the membership in the IPR category. To obtain a variance for the estimate of the number insured, independence must be assumed or the covariance needs to be modeled. The current county model assumes independence and thus may misrepresent the true variance of the estimate.

The fit of the IPR estimates is good overall with some exceptions; there were no indications that the model is misspecified with respect to the logistic regression function, and the distributional assumptions fit well most of the time, with some exceptions for IPR proportions close to zero or one. However, the county model does not do as well in predicting health insurance coverage for low-income women ages 50-64. For uninsured women between the ages of 40-64 and with incomes less than or equal to 200 percent of the federal poverty threshold, the CV is .36.

These results point to an area in need of further research. The posterior distributions are sensitive to the variance model and the variance models seem mostly consistent with the data and with the 'official' variance estimates (14). The distributional assumptions, including those for normality and for the sampling error variances, are important assumptions and should be examined further. Thus, the posterior distributions are preliminary.

Future Directions

We have established the feasibility for producing estimates of health insurance coverage for the states and counties. The current version of our model of the NBCCEDP eligible population has produced estimates by age, race, ethnicity, and IPR for states. We have also produced estimates by age, sex and IPR for counties, but some counties may need to be aggregated into larger areas if there is a future public release file. However, many

simplifying assumptions have been made. As our research continues, these assumptions will be examined, new categories of race and ethnicity will be incorporated, and new data will be added. We expect the precision of the estimates to improve. It remains to be decided if the current level of precision is adequate for program planning and evaluation.

References

1. Ehemann C.R., Benard V.B., Blackman D., et al. Breast cancer screening among low income or uninsured women: results from the National Breast and Cervical Cancer Early Detection Program, July 1995-March 2002. *Cancer Causes and Control*. 2006; 17:29-38
2. Swan J., Breen N., Coates R.J., Rimer B.K., Lee N.C. Progress in cancer screening practices in the United States: results from the 2000 National Health Interview Survey. *Cancer*. 2003; 97(6):1528-1540.
3. US Preventive Services Task Force. *Guide to Clinical Preventive Services, 3rd Edition: Periodic Updates*. October 2002.
<http://www.ahrq.gov/clinic/3rduspstf/preface.htm>.
4. <http://www.cdc.gov/cancer/nbccedp/info-bc.htm>
5. <http://www.cdc.gov/std/HPV/ScreeningTables.pdf>
6. http://www.cancer.org/docroot/CRI/content/CRI_2_4_2X_What_are_the_risk_factors_for_breast_cancer_5.asp
7. Ayanian J.Z., Kohler B.A., Abe T., and Epstein A. “The Relationship Between Health Insurance Coverage and Clinical Outcomes Among Women with Breast Cancer”. *New England Journal of Medicine*. 1993; 329(5): pp. 326-331.
8. Ryerson A.B., Benard V.B., and Major A.C. The National Breast and Cervical Cancer Early Detection Program 1991–2002 National Report. Atlanta (GA): Centers for Disease Control and Prevention, Department of Health and Human Services; 2005.
9. <http://www.cdc.gov/cancer/nbccedp/about2004.htm>
10. <http://www.census.gov/hhes/www/sahie/index.html>
11. Gilks, W.R., Richardson, S., and Spiegelhalter, D., eds. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall. 1996.
12. Prasad, N.G.N. and Rao, J.N.K. “The Estimation of the Mean Squared Error of Small-Area Estimators”, *Journal of the American Statistical Association*. 1990; 85: 163-171.
13. Gelman, A. and Meng, X. Model Checking and Model Improvement. In *Markov Chain Monte Carlo In Practice*. Edited by W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. 1996; pp. 189-201.
14. DeNavas-Walt, C., Proctor, D. and Hill Lee, C. *Income, Poverty, and Health Insurance Coverage in the United States: 2004*. U.S. Census Bureau, Current Population Reports, P60-229. Washington, DC: U.S. Government Printing Office. 2005.