Using Census and Mortality Data to Target Small Areas for Breast, Colorectal, and Cervical Cancer Screening

ABSTRACT

Objectives. The goal of this study was to develop and validate quantitative models for estimating cancer incidence in small areas.

Methods. The outcome for each cancer site was the incidence of disease that had reached a late stage at the time of diagnosis. Two sets of predictors were used: (1) census-based demographic variables and (2) census-based demographic variables together with the cancer-specific mortality rate.

Results. The best models accounted for a substantial percentage of between-area variability in late-stage rates for cancer of the breast (46%), cervix (61%), and colon/rectum (58%). A validation procedure indicated that correct identification of small areas with high rates of late-stage disease was two to three times more likely when model-based estimates were used than when areas were selected at random.

Conclusions. Additional testing is needed to establish the generality of the geographic targeting methodology developed in this paper. However, there are strong indications that small-area estimation models will be useful in many regions where planners wish to target cancer screening programs on a geographic basis. (*Am J Public Health.* 1994;84:56–61)

Howard F. Andrews, PhD, Jon F. Kerner, PhD, Ann G. Zauber, PhD, Jeanne Mandelblatt, MD, MPH, John Pittman, MS, and Elmer Struening, PhD

Introduction

Models for cancer screening programs typically focus on subgroups of the general population on the basis of age- and sex-specific incidence, prevalence, and mortality rates of a particular cancer. Although analyses of individual characteristics indicate who should be screened, small-area analysis can be used to determine where to set up screening programs so that high-risk populations can be more easily recruited. In previous studies of small-area variation in cancer incidence, we quantified the benefits of geographically targeting cancer screening programs, using clustered census tracts as the unit of analysis.[1,2] Those analyses, based on cancer registry data, indicated that the efficiency of breast cancer screening programs can be substantially improved by targeting communities with known high incidence rates.

In many localities, there is no cancer registry to provide incidence data at the community level, so targeting based on known small-area incidence, using the methodology described above, is not possible. Here we describe a two-step technique for making geographic targeting available to these localities:

1. Develop a small-area multiple regression model relating incidence for a particular cancer site to mortality, census-based variables, or both in a locality where cancer registry data are available.
2. Apply the model in the locality of interest, where cancer registry data are not available, and obtain an estimated incidence value for each small area.

The outcome in the regression models is the incidence rate of disease staged as regional or distant at the time of diagnosis (late-stage incidence). The main goal of screening programs is to reduce this rate by identifying in situ and localized cancer in individuals who would otherwise have developed late-stage disease before diagnosis and treatment.

Methods

Geocoding Cancer Incidence and Mortality

New York City cancer incidence data for the years 1976 through 1982 were obtained from the New York State Department of Health Cancer Registry. Mortality data for the years 1977 through 1983 were obtained from the New York City Department of Health. The geographic coding of these data has been previously described.[1] Denominators for small area rates were obtained from 1980 US census data (Summary Tape File 3a).

Incidence, mortality, and census data were aggregated to health areas, administrative units used by the New York City Department of Health. Each health area, with an average population of about 21 000, consists of four to six census tracts. Data were obtained for four of the five boroughs that make up New York City: Bronx, Brooklyn, Manhattan, and

Howard F. Andrews, John Pittman, and Elmer Struening are with the New York State Psychiatric Institute, New York, NY. At the time of the study, Jon F. Kerner, Ann G. Zauber, and Jeanne Mandelblatt were with the Memorial Sloan-Kettering Cancer Center, New York, NY. Dr Kerner is now with the Vince T. Lombardi Cancer Research Center, Georgetown University Medical Center, Washington, DC.

Requests for reprints should be sent to Howard F. Andrews, PhD, Epidemiology Department, New York State Psychiatric Institute, 722 W 168th St, Box 47, New York, NY 10032.

This paper was accepted May 11, 1993.

Queens. To reduce rate instability due to small denominators, health areas with a total 1980 population of less than 5000 were excluded from the analysis. Areas in which more than 10% of the total population resided in an institution were also excluded. The 22 health areas excluded by these criteria contained 4.2% of the total population of the four boroughs. The 319 health areas that were included contained a total population of 6 637 686 in 1980.

Rate Calculation

It is customary to adjust for age when comparing geographic variability in cancer rates. However, when targeting geographic areas for screening service delivery, a concentration of age groups at higher risk for a particular cancer should be considered as part of the planning process. We therefore calculated crude age-truncated rates for each area by dividing the total number of incident late-stage cases at or above a specific age by the total area population of the appropriate sex at or above the same age. For cervical cancer, we used the incidence of invasive disease as the measure of late stage. For breast and colorectal cancer, late stage was defined according to the Cancer Registry categories of regional or distant disease.

For each cancer site, an age-truncated crude late-stage incidence rate and an age-truncated crude mortality rate were calculated. For breast cancer, incidence and mortality rates were calculated for women aged 35 years and older, for cervical cancer, rates were calculated for women aged 20 years and older; for colorectal cancer, we combined men and women aged 45 years and older in calculating rates. No upper age limit was set for any site. These age-truncated rates include more than 95% of registry cases for each of the three sites. The total number of late-stage cases used for rate calculation was 11 634 for breast cancer, 16 140 for colorectal cancer, and 3640 for cervical cancer. For mortality rate calculation, the numerator for each cancer site represents all individuals for whom the primary cause of death was cancer of that site during the period 1977 through 1983: 9132 for breast cancer, 9212 for colorectal cancer, and 1285 for cervical cancer. Incidence data and mortality data were not linked: the numerator for each mortality rate includes all deaths for that cancer site in the stated time period, not just deaths of individuals reported to the Cancer Registry during the period for which incidence rates were calculated.

All rates are expressed in terms of yearly averages per 100 000 population. The formula for rate calculation is

$$[(N/7)/T] \cdot 100\,000,$$

where N is the total number of late-stage cases in a given health area for a particular cancer site among individuals in the designated age group during the 7-year study period and T is the total 1980 population in the health area for the site-defined age and sex group. Hereinafter, age-truncated late-stage incidence rates for breast and colorectal cancer and the rate of invasive cervical cancer will be referred to as late-stage rates.

Selection of Predictors for Use in Models to Estimate Late-Stage Incidence Rates

In choosing census-based predictors, we were guided by two criteria: known etiologic relationship to one or more of the three cancers under consideration and absence of multicollinearity among predictors. We were aided in variable selection by previous research that identified dimensions of demography and housing characteristics in New York City.[3] Beginning with the 14 census-based variables that defined these dimensions, we used multiple regression analysis to isolate a subset of four variables that met the above criteria and that, when used together, accounted for nearly as much variability (adjusted R^2) in each of the late-stage rates as the full set of 14 variables: (1) percentage of the population aged 65 years and older; (2) percentage of households with income higher than $50 000 per year; (3) percentage of the population (aged 15 years and older) divorced or separated; and (4) percentage of women in the labor force with one or more children aged 16 years or younger. Descriptive statistics for these variables, together with mortality rates and late-stage incidence rates, are shown in Table 1. For analytic purposes, we applied log or square root transformations to skewed variables to approximate a normal distribution, as indicated in Table 1.

Data from all four boroughs were used to develop a multiple regression model for each of the three late-stage rates, with the four census variables and the appropriate mortality rate as predictors. To simulate a situation in which targeting is to be done in a locality for which cause-specific mortality data are not available or not accessible, we also developed a model for each cancer in which only census variables were used as predictors. The addition of interaction terms did not improve the adjusted R^2 values for any of the regression models.

TABLE 1—Descriptive Statistics for Outcome and Predictor Variables (n = 319 Health Areas)

	Mean	SD
Outcome variables—incidence rates[a]		
Late-stage breast cancer	91.72	23.92
Invasive cervical cancer[b]	22.44	12.76
Late-stage colorectal cancer	94.48	24.58
Predictors—mortality rates[a]		
Breast cancer[b]	71.66	21.85
Cervical cancer[b]	8.22	6.68
Colorectal cancer	79.22	21.32
Predictors—census variables		
% persons aged 15+ y divorced or separated	11.98	4.81
% working mothers with children <18 y	29.48	10.43
% households with income above $50 000[c]	3.30	4.76
% persons aged 65 y and older	13.16	5.70

[a]Expressed as cases per 100 000 population.
[b]A square root transformation was applied to this variable in the regression models.
[c]A log transformation was applied to this variable in the regression models.

Model Validation

For each cancer site, we used the following procedure to determine the accuracy with which a model developed in one geographical area could predict late-stage rates in another area.

1. One of the four boroughs of New York City—for example, Queens—was designated as the geographic area for which estimated rates of late-stage disease for a particular site—say breast cancer—were needed.

2. A multiple regression model for estimating late-stage breast cancer rates was developed, using only data from the other three boroughs—in this case, Bronx, Manhattan, and Brooklyn.

3. The regression model was used to calculate an estimated late-stage rate for each health area in Queens. This was done by taking the known value of each predictor for a particular health area in Queens and multiplying by the appropriate unstandardized regression coefficient specified in the three-borough model. These

TABLE 2—Correlation[a] between Estimated and Actual Late-Stage Incidence Rates, Based on Models with Mortality and Census Variables and Census Variables Only

Borough	Census and Mortality	Census Only
Breast cancer		
Brooklyn	.69	.62
Bronx	.74	.70
Manhattan	.63	.46
Queens	.53	.40
Mean	.65	.55
Cervical cancer		
Brooklyn	.78	.78
Bronx	.73	.63
Manhattan	.79	.72
Queens	.74	.67
Mean	.76	.70
Colorectal cancer		
Brooklyn	.77	.60
Bronx	.80	.83
Manhattan	.68	.44
Queens	.68	.50
Mean	.73	.59

[a]Pearson's product-moment correlation coefficient (*r*).

TABLE 3—Accuracy of Targeting on the Basis of Estimation Models

	Model Type			
	Census and Mortality		Census Only	
Borough	Top 20%	Top 40%	Top 20%	Top 40%
Breast cancer				
Brooklyn	45	81	32	64
Bronx	73	82	73	82
Manhattan	47	74	40	67
Queens	47	73	40	53
Mean	53	76	46	67
Cervical cancer				
Brooklyn	59	91	41	82
Bronx	64	82	55	82
Manhattan	73	100	53	93
Queens	53	80	53	86
Mean	62	88	51	86
Colorectal cancer				
Brooklyn	68	82	50	77
Bronx	64	100	82	91
Manhattan	67	87	53	60
Queens	60	73	53	80
Mean	65	86	60	77

Note. Numbers in table are percentages of targeted health areas (areas with estimated late-stage rates in the top 20%) that fell into the top 20% or top 40% of all health areas with respect to actual late-stage incidence rates. Higher percentages indicate greater accuracy.

terms were then summed and the value of the model intercept was added, yielding the estimated late-stage rate for that health area. This process was repeated for each health area in Queens.

4. The model was validated by correlating the estimated late-stage rates for the 74 health areas of Queens, based on the model developed with data from the other three boroughs, with the actual late-stage rates in those same areas, based on data from the New York State Department of Health Cancer Registry.

For each cancer site, the above steps were repeated. Data from each unique combination of three boroughs were used to build the model and obtain an estimated rate of late-stage disease for each health area in the excluded borough. For each site and borough, one model was built with the four census-based variables and the site-specific mortality rate as predictors; the other model included only the four census-based predictors.

Results

The correlation between actual late-stage rates and late-stage rates estimated from the appropriate model is shown for each combination of borough and cancer site in Table 2. In general, the association between estimated and actual rates was highest for invasive cervical cancer, for which all correlations were greater than .7 when mortality and census data were used in the estimation model. As shown in the table, all correlations were greater than .6 even when census data alone were used. The lowest correlations were obtained for the breast cancer models: the mean correlation was .65 when mortality was included and .55 when only census variables were used. For colorectal cancer, the correlation between estimated and actual late-stage rates varied from a low of .68 in Manhattan to a high of .80 in the Bronx when mortality and census data were used. When only census data were used in the model, correlations ranged from .44 in Manhattan to .83 in the Bronx.

From a planner's perspective the key question with respect to model validity is, What is the relationship between actual and estimated rates *in the areas targeted to receive a particular service?* To answer this question, we follow a hypothetical scenario in which planners have decided to target for screening the health areas with the highest estimated late-stage rates, defined as those areas in the top quintile (20%).

To determine the accuracy of this targeting process, we calculated the percentage of targeted areas in each borough that fell in (1) the upper 20% of health areas with respect to *actual* late-stage rates and (2) the upper 40% of health areas with respect to *actual* late-stage rates. If the model provided perfect predictions, 100% of the targeted areas would be in the upper 20% with respect to actual late-stage rates. In the worst-case scenario, in which the model had no predictive power, one would expect that on average only 20% of the targeted areas would fall in the upper quintile of actual late-stage rates—the proportion expected by chance. Similarly, with perfect prediction, 100% of the targeted areas would be in the upper 40% of areas with respect to actual late-stage rates; in the worst case, only 40% of the targeted areas would be in the upper two quintiles of actual late-stage rates. The normal approximation to the binomial distribution was used to determine whether the actual number of targeted areas in the upper quintile was significantly greater than the number expected by chance. The results are shown in Table 3.

The concordance between targeted areas and the upper quintile of actual late-stage rates is greatest for invasive cervical and colorectal cancer and lowest for breast cancer. For all cancer sites, the concordance between upper quintile of estimated and actual rates is considerably better than what would be expected by chance ($P < .00001$). For example, as indicated in the first row, first column of Table 3, 45% of the areas in the upper quintile of estimated late-stage rates of breast cancer in Brooklyn (the targeted areas) were also in the upper quintile of actual late-stage rates—more than twice the chance expectation of 20%. As indicated in the first row, second column, 81% of the health areas that were in the upper quintile with respect to estimated late-stage rates were either in the highest or second highest quintile in terms of actual late-stage rates, more than twice the 40% rate of concordance expected by chance. Consistent with the correlational analyses (Table 2), concordance between estimated and actual rates is better in models using mortality and census-based predictors than in models using only census-based predictors.

Maps of the results with respect to one of the outcomes, invasive cervical cancer, are shown for the Bronx (Figure 1), providing visual confirmation of the strong relationship between estimated and actual rates. The maps also show that the high-rate areas that would be targeted for screening in the Bronx tend to cluster geographically. Such clustering should enhance the efficiency with which screening resources can be geographically allocated.

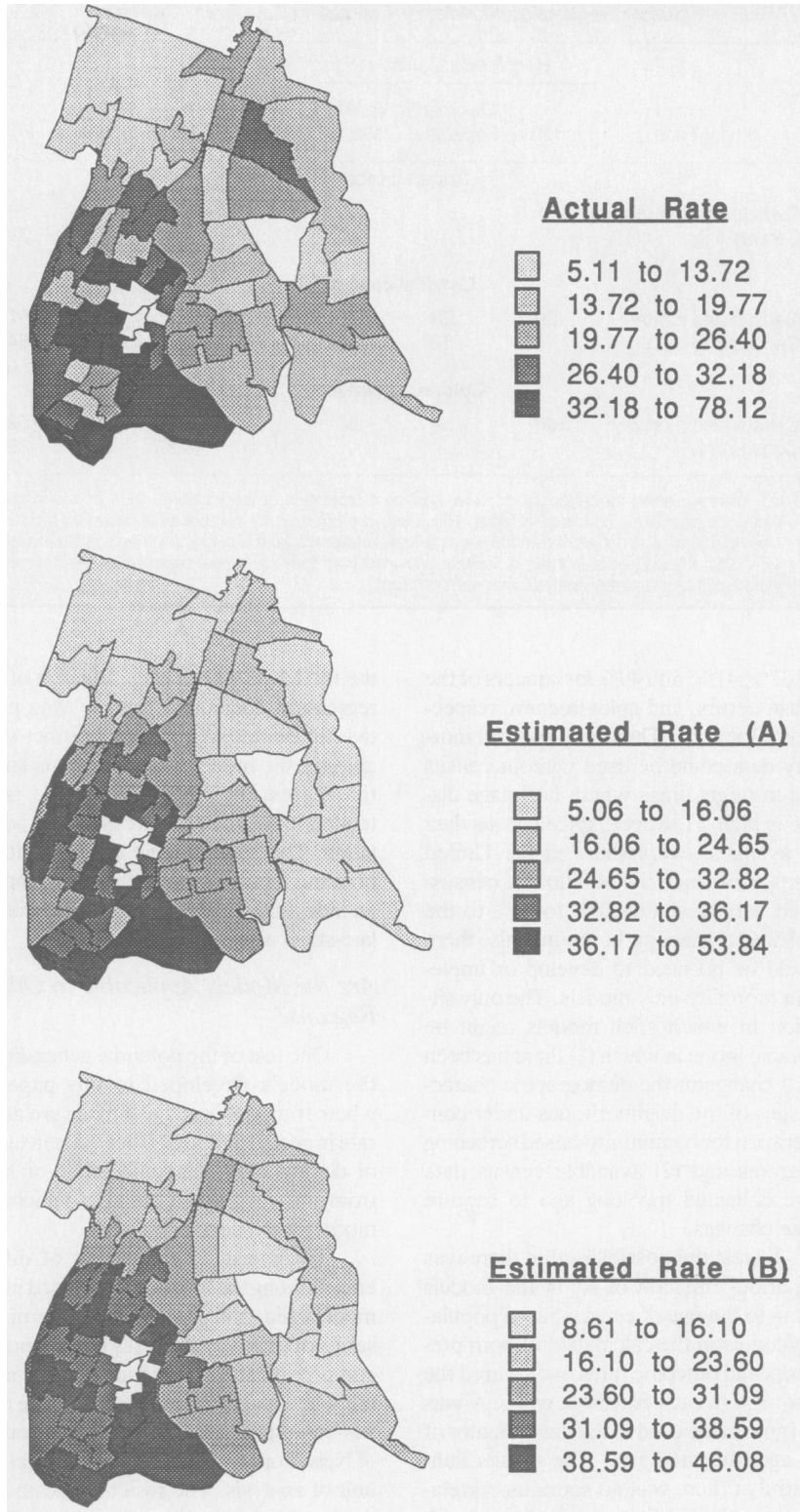
"Best" Models

In the validation process described above, several models were developed for each cancer site, each based on data from three boroughs. Given the apparent viability of targeting small areas on the basis of census data alone or census and mortality data, the question arises, What specific model should be used in applying this approach in a city or region in which cancer registry data are not available?

To develop a set of "best" models for use by planners, we used data from all four boroughs to maximize the stability of parameter estimates. As shown in Table 4, the four census variables alone explained a statistically significant proportion of variability in each of the late-stage rates ($P < .001$): the adjusted R^2 was .33 for breast cancer, .54 for invasive cervical cancer, and .39 for colorectal cancer. In the models using breast cancer and colorectal cancer as outcomes, only one census variable accounted for a significant proportion of variability in late-stage rates—the percentage of the population in the health area that was 65 and older. For cervical cancer, two census variables were significantly associated with the rate of invasive disease: (1) percentage of the population divorced or separated and (2) income level—poor areas had significantly higher rates than wealthier areas.

For each cancer site, the addition of the site-specific mortality rate to the model added significantly to the proportion of explained variability. The greatest increment in adjusted R^2 resulting from the addition of mortality rate as a predictor occurred for colorectal cancer (19%). Moreover, for this cancer site two census variables that were not significant in the census-only model, percentage of working mothers and percentage of population divorced or separated, did make significant contributions to R^2 when the mortality rate was added.

When the site-specific mortality rate was used without census variables in the model, a substantial proportion of variability in late-stage rates was accounted



Note. Areas with rates in the upper quintile (20%) of all health areas—those that would be targeted in a screening program for cervical cancer—are shown in the darkest shading. Unshaded (white) areas had populations that did not meet criteria for inclusion in the analysis. Estimated rates, based on models developed using data from health areas in Brooklyn, Manhattan, and Queens, are shown in the two lower maps. In model A, estimated rates were based on regression analysis in which both mortality rates and census variables were used as predictors. In model B, rates were based on an analysis in which only census variables were used.

FIGURE 1—Geographical distribution of actual and estimated rates of invasive cervical cancer in health areas in the Bronx (n = 57).

TABLE 4—Multiple Regression Models for Estimating Late-Stage Incidence

Model Type	Regression Coefficient (Unstandardized) Mortality	Divorced/ Separated	Working Mother	High Income	Elderly	Intercept	R^2
Breast cancer							
Census and mortality	7.01	...	...	...	10.46	−3.79	.46
Census only		...	...	...	12.47	30.72	.33
Cervical cancer							
Census and mortality	.39	.09	...	−.52	...	3.00	.61
Census only		.13	...	−.67	...	3.75	.54
Colorectal cancer							
Census and mortality	.60	−.58	.30	...	10.86	6.57	.58
Census only		...	...	...	19.33	26.07	.39

Note. The regression model accounts for a significant proportion of the variability (R^2) in late-stage incidence rate in all models ($P < .001$). R^2 values are adjusted for number of variables in model. Coefficients of all variables shown have a significant relationship to the outcome variable in the model ($P < .05$). Ellipsis points indicate a variable excluded from the final model because its effect when included in a preliminary analysis was nonsignificant.

TABLE 5—Multiple Regression Models for Estimating Late-Stage Incidence Based Only on Mortality

Cancer	Mortality Regression Coefficient	Intercept	R^2
Breast	10.03	8.07	.37
Cervical	.75	2.58	.41
Colorectal	.81	30.24	.49

Note. The regression model accounts for a significant proportion of the variability (R^2) in late-stage incidence rate in all models ($P < .001$). R^2 values are adjusted for number of variables in model.

for: 37%, 41%, and 49% for cancers of the breast, cervix, and colon/rectum, respectively (Table 5). Thus, site-specific mortality data could be used without census data to target areas where late-stage disease is high. However, since census data are available everywhere in the United States, and since the addition of census-based variables adds 10% to 20% to the explained variability in the models, there should be no need to develop or implement mortality-only models. The only situation in which such models might be valuable is one in which (1) there has been rapid change in the demographic characteristics of the neighborhoods under consideration for community-based screening programs and (2) available census data were collected too long ago to capture these changes.

To test the possibility that there was a spurious inflation of R^2 in the models owing to the use of census-based population counts in the calculation of both predictors and outcome rates, we entered the factor $1/N$ in each equation, where N was the same value used in the denominator of the age-truncated rate. The results indicated that there was no spurious correlation: in each model the significance of each predictor was unaffected by the addition of $1/N$, and the values of the regression coefficients remained approximately the same as those shown in Table 4. However, for two of the cancer sites, breast and cervix, the term $1/N$ was itself significant in the model. In the case of late-stage breast cancer, the adjusted R^2 was increased by 2% by the addition of $1/N$ to the model. For both sites, the sign of the regression coefficient for $1/N$ was positive, indicating that when the other variables in the model were held constant in the regression model, late-stage rates tended to be higher in areas of low population. This suggests that low population, or some factor associated with it, makes an independent contribution to the risk of late-stage disease.

Are the Models Applicable to Other Regions?

One test of the potential generality of the models developed in this paper is whether model-based estimates are accurate in geographic areas that differ in terms of demography, outcome rates, or both from the geographic areas in which the models were developed.

To determine the extent of differences among the four boroughs used in the model validation process, we performed a series of one-way analyses of variance to compare mean values of site-specific mortality late-stage rates and each of the census-based predictors in the four boroughs of New York City, using health area as the unit of analysis. The results indicate that the four boroughs differed substantially in terms of key demographic and outcome characteristics. All pairs of boroughs except Manhattan and Brooklyn differed significantly in terms of mean values of at least one late-stage rate. Four of the six pairs differed in terms of at least one site-specific mortality rate, and all six pairs differed in terms of mean value of at least one demographic characteristic.

Discussion

The results indicate that a quantitative model developed in one geographically defined area on the basis of available census data can be used to obtain good estimates of late-stage rates for breast, colon/rectum, and cervical cancer in another geographic area. The accuracy of estimation is improved for all three cancer sites when the site-specific mortality rate is included as a predictor. Because disease that has progressed to a late stage at the time of detection is likely to result in death, it is not surprising that mortality data considerably enhance the power of the model to estimate late-stage rates. More unexpected is the finding that without mortality rate as a predictor, and using only four variables derived from census data, the models can account for a substantial proportion of geographic variability in late-stage rates of all three cancer sites. Therefore, even in localities without ready access to population-based mortality data, it should be possible to estimate late-stage rates accurately and to successfully target screening programs to high-risk communities. However, all of the areas used in developing and testing of the models presented in this paper are located in New York City. Ultimately, these models will be useful only if they can be applied in targeting community-based screening programs in other regions of the United States and the world. Thorough testing will require comparison of model-based estimates with actual rates in other regions of the country where cancer registry data are available.

There are a number of indications that these models will work well in other geographical regions. First, the results demonstrate that the four boroughs of New York City differ substantially along

the dimensions of demography and cancer outcome examined in this paper. This indicates that the models are robust with respect to demographic and disease characteristics of the region for which estimates of late-stage cancer are required.

An examination of the predictors used to estimate late-stage rates also suggests that the models will be widely applicable. Site-specific mortality rates would be expected to be strongly associated with late-stage incidence in any community. For breast and colorectal cancer, an increase in overall incidence as well as late-stage disease with advancing age has been well documented,[4,5] so the percentage of an area's population that is 65 years old or older would be expected to predict late-stage rates for these two sites in any locality. Similarly, invasive cervical cancer and late-stage colorectal cancer are known to be more prevalent in poor populations,[4,6] so the association of low income and socioeconomic factors with late-stage incidence that is quantified in the models developed in New York City should apply in other areas as well.

However, the relationship between predictors and cancer outcomes will no doubt be attenuated if the models are applied to regions in which there is little variability, as would be the case in, say, a region made up of middle-income suburban communities. The models developed in this paper are therefore most appropriate in areas that contain substantial demographic variability among the communities under consideration for targeting.

The finding that performance of the models was poorest for breast cancer is consistent with the results of a previous analysis,[2] which indicated that of the three cancer sites examined, the year-to-year reliability of late-stage rates was lowest for breast cancer. When the reliability of any outcome measure is relatively low, the ability to account for variability in the outcome on the basis of other measures will be limited.[7] However, the results indicate that even for breast cancer, the models correctly identified high-risk geographic areas at considerably better than twice the level that would be expected by chance.

Methods used in some small-area analyses involving health outcomes have been criticized for failing to adequately test the null hypothesis that observed differences between late-stage rates are due to random variation.[8,9] When data from multiple years are available, the method we have developed to quantify year-to-year rate stability[2] and that was applied to the outcomes examined in this paper can be used to directly test the null hypothesis of random between-area variation.

Can the estimation models be improved? Some of the unexplained variability in each model may reflect measurement error or factors that are specific to particular geographic regions. However, it may be possible to improve the R^2 value of the models by including additional demographic factors. The finding that low population is related to high rates of late-stage disease, after controlling for sociodemographic factors, suggests one direction for future research. In New York City, health area boundaries were originally defined in such a way that each area had a total population of approximately 20 000. Health areas with low population have resulted from loss of housing in certain areas.[10] The impact of loss of housing and social disintegration on the health status of individuals who remain in these areas is well documented.[11] It may be that census-based measures of housing loss or other indicators of deterioration over time will improve the ability of estimation models to identify appropriate sites for targeting community-based cancer screening and other health-related intervention programs.

In conclusion, there is good reason to believe that the relationships among late-stage cancer incidence, mortality, and demographic characteristics that have been quantified in the models presented in this paper will apply in many urban and nonurban regions. If this hypothesis can be confirmed in future research, geographically based targeting of cancer screening programs will no longer be limited to areas in which population-based cancer registries have been established. □

Acknowledgments

This investigation was supported in part by NCI grant J1 R01 CA 46586.

We would like to express our thanks to Philip C. Nasca, PhD, Director, Bureau of Cancer Epidemiology, New York State Department of Health, and to William Burnett, MD, Director, New York State Cancer Incidence Registry, who made Registry information available and generously shared their expertise in the analysis of these data. We thank Drs Patricia Cohen, Bruce Link, and Rodrick Wallace and two anonymous reviewers for their valuable comments on earlier versions of the manuscript.

References

1. Kerner JF, Struening E, Pittman J, Andrews H, Sampson N, Strickman N. Small area variation in cancer incidence and mortality: a methodology for targeting cancer control programs. In: Engstrom P, Anderson P, Mortenson L, eds. *Advances in Cancer Control: Epidemiology and Research.* New York, NY: Alan R. Liss Inc; 1984:225–234.
2. Kerner J, Andrews H, Zauber A. Geographically-based cancer control: methods for targeting and evaluating the impact of interventions on defined populations. *J Clin Epidemiol.* 1988;41:543–553.
3. Pittman J, Andrews H, Struening E. The use of zip coded populations in social area studies of service utilization. *Eval Program Plann.* 1988;9:309–317.
4. Mandelblatt J, Andrews H, Kerner J, Zauber A, Burnett W. Determinants of late stage diagnosis of breast and cervical cancer: the impact of age, race, social class, and hospital type. *Am J Public Health.* 1991;81:646–649.
5. Ries L, Henkey B, Edwards B, eds. *Cancer Statistics Review.* Bethesda, Md: National Institutes of Health; 1990. Publication 90-2789.
6. Baquet CR, Horm JW, Gibbs T, Greenwald P. Socioeconomic factors and cancer incidence among blacks and whites. *J Natl Cancer Inst.* 1991;83:551–557.
7. Fleiss JL. *The Design and Analysis of Clinical Experiments.* New York, NY: John Wiley & Sons Inc; 1986.
8. Diehr P, Cain K, Connell F, Volinn E. What is too much variation? The null hypothesis in small area analysis. *Health Serv Res.* 1990;24:741–771.
9. Diehr P. A small area simulation approach to determining excess variation in dental procedure rates. *Am J Public Health.* 1990; 80:1343–1348.
10. Struening EL, Wallace R, Moore R. Housing conditions and the quality of children at birth. *Bull N Y Acad Med.* 1990;66:463–478.
11. Wallace RW, Wallace D. Origins of public health collapse in New York City: the dynamics of planned shrinkage, contagious urban decay and social disintegration. *Bull N Y Acad Med.* 1990;66:391–434.