

## ORIGINAL RESEARCH

# Spatial Analysis of Breast Cancer Mortality Rates in a Rural State

Marisa Schulz, MSc<sup>1</sup>; Emma Spors, MSc<sup>1</sup>; Kari Bates<sup>2</sup>; Semhar Michael, PhD<sup>1</sup>

Accessible Version: [www.cdc.gov/pcd/issues/2022/22\\_0113.htm](http://www.cdc.gov/pcd/issues/2022/22_0113.htm)

*Suggested citation for this article:* Schulz M, Spors E, Bates K, Michael S. Spatial Analysis of Breast Cancer Mortality Rates in a Rural State. *Prev Chronic Dis* 2022;19:220113. DOI: <https://doi.org/10.5888/pcd19.220113>.

## PEER REVIEWED

**Summary****What is known on this topic?**

Breast cancer affects 1 in 8 women in the US and is the most frequently diagnosed cancer in women.

**What is added by this report?**

Because South Dakota is a rural state, sociodemographic factors affect the population differently than in the general US population. We assessed the spatial distribution of breast cancer mortality rates by county, and our findings add insight on educational attainment as a risk factor for breast cancer.

**What are the implications for public health practice?**

Our results can be used to help allocate resources to the South Dakota counties that need them most.

## Abstract

**Introduction**

Breast cancer affects 1 in 8 women in the US and is the most frequently diagnosed cancer in women. In South Dakota, 102 women die from breast cancer each year. We assessed which sociodemographic factors contributed to mortality rates in South Dakota and used spatial analysis to investigate how counties' observed age-adjusted mortality rates compared with expected rates.

**Methods**

We computed standardized incidence ratios (SIRs) of all counties in South Dakota by using the age-adjusted mortality rates, the 2000 US standard population, and the South Dakota estimated population. We used a linear regression model to identify sociodemographic factors associated with breast cancer mortality

rates and to compute a new SIR value, after controlling for relevant factors.

**Results**

Educational level and breast cancer incidence rates were significantly associated with breast cancer mortality rates at the county level. The SIR values based on age-adjusted counts showed which counties had more deaths due to breast cancer than what might be expected using South Dakota as the reference population. After controlling for sociodemographic factors, the range of SIR values decreased and had lower variability.

**Conclusion**

The regression model helped identify factors associated with mortality and provided insights into which risk factors are at play in South Dakota. This information, in combination with the spatial distribution of mortality by county, can be used to help allocate resources to the counties in South Dakota that need them most.

## Introduction

In South Dakota, breast cancer is the most commonly diagnosed cancer and the second leading cause of cancer death among women (1,2). In 2022, an estimated 750 new cases and 110 deaths attributed to female breast cancer will occur in South Dakota. In general, a woman in the US has a 1-in-8 lifetime risk of developing breast cancer (3,4). Since 1989, the US breast cancer mortality rate has decreased 40%, but from 2010 to 2019 the rate slowed to a low of decreasing by 1.3% per year (5).

Characteristics such as age and race and ethnicity affect a woman's chances of being diagnosed with or dying of breast cancer, but new evidence has established that sociodemographic factors, including education level, also play a role (6). Albano et al noted a negative relationship between number of years of education and cancer mortality and found that the level of education and race vary considerably with mortality rates (7). Of the South Dakota population aged 25 years or older, 92.2% are high school graduates (higher than the national average) and 29.3% have a bachelor's degree or higher (lower than national average) (8).



The opinions expressed by authors contributing to this journal do not necessarily reflect the opinions of the U.S. Department of Health and Human Services, the Public Health Service, the Centers for Disease Control and Prevention, or the authors' affiliated institutions.

Olson et al acknowledged that communities exist in which geographic disparities are more prominent because of rural isolation and small population size (9). Furthermore, 64 of 66 counties in South Dakota are categorized as rural or frontier, and South Dakota contains 9 American Indian reservations (10). Finally, 61.6% of women receiving breast services are White, and 16.7% are American Indian (11); most of the population in South Dakota is White, and the leading minority is American Indian at 8.8% (8).

The study aimed to describe the spatial distribution of female breast cancer mortality at the county level in South Dakota and assess the association between mortality rates and risk factors reported in the literature.

## Methods

### Data source

The 66 counties of South Dakota have boundaries that are defined by the South Dakota Legislature and accepted by the US Census. The counties range in population from 183,439 in Minnehaha County to 917 in Jones County, and the median population per county is 5,413. Most residents of South Dakota were White; the median percentage of non-White residents by county was 6.6% and the maximum was 95.2%

### Cancer data

Breast cancer incidence and mortality rates from 2001 through 2015 were extracted from the South Dakota State Cancer Registry provided by the South Dakota Department of Health (12). Both rates were per 100,000 people and age-adjusted to the 2000 US standard population and the South Dakota estimated population. The proportion of mammography screening rates in South Dakota was based on the numbers reported by Holzhauser et al for the All Women Count! mammography program (13). The average number of participants for 1997 through 2016 was reported for the program; then, the average number of participants was adjusted for the total number of women older than 40 years in the county to get an estimated screening rate for each county (13,14).

### Demographics

We used 2015 data from the US Census Bureau to obtain information on the 66 South Dakota counties, including the number of providers and the education level, poverty level, percentage of uninsured, median age, and race of residents (8).

Data on educational attainment were obtained from the US Census Bureau's American Community Survey (ACS). These data were count estimates for the population of each county aged 25 years or older. Levels were categorized as less than 9th grade, 9th through

12th grade with no diploma, high school graduate or equivalent, some college but no degree, associate degree, bachelor's degree, and graduate or professional degree. These values were modified into an educational attainment statistic of the percentage of the population with less than a bachelor's degree of education. The statistic used in this study was the percentage of the population with less than a bachelor's degree, by county.

We collected data on poverty estimates, by county, from the ACS; these data adhered to the standards specified by the Office of Management and Budget in Statistical Policy Directive 14 (8). Poverty was determined by a set of income thresholds that consider the living situation (alone or with nonrelatives), age, and number of people per household. For example, the poverty threshold for 2-person families varies by the age of the primary householder and differs from the poverty threshold for people living alone or with nonrelatives, which also varies by age.

Insurance coverage percentages were collected from Small Area Health Insurance Estimates (SAHIE) (15). The uninsured percentage included residents who were not covered by insurance, which excluded those on government assistance such as Medicaid or Medicare. Finally, the data set summarizing racial distributions in a county included estimated population counts for American Indian and Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, other race, and 2 or more races. Because of South Dakota's predominantly White population, the data were configured into White and non-White, which determined the non-White percentage per county (8) (Table 1).

### Statistical analysis

#### Data manipulation and missing value imputation

Mortality rates and the various independent variables were combined into 1 data set; 15 of the 66 counties were missing mortality rate data. Mortality rates are often suppressed from public availability when 3 or fewer deaths are reported in a county, to protect patient identity. To remedy the missing data, *k*-nearest-neighbor (KNN) imputation was used to estimate the missing mortality values. KNN imputation compares a data point  $x_i$  with its *k* nearest neighbors and then approximates  $x_i$  using the majority vote of these *k* neighbors in multidimensional space. For the data, *k* = 9 nearest neighbors were used, and a weighted mean of the *k* nearest values was placed for each missing  $x_i$  (16,17). This was done with the function "knn()" from the R package VIM version 6.1.1 (R Foundation for Statistical Computing) (18).

#### Multiple linear regression

We used multiple linear regression to model the relationship between the factors in this study and breast cancer mortality rates

(19). We considered several potential predictor variables with observed correlation, hence a stepwise variable selection technique was used, in both the forward and backward direction, to perform feature selection. As a result, a subset of the factors that were associated with the mortality rates was obtained based on Akaike information criterion (AIC) (19,20).

The resulting model is of the form  $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_p X_{pi}$  for  $i = 1, 2, \dots, 66$  where  $\hat{Y}_i$  represents the estimated mortality rate for the  $i^{\text{th}}$  county and  $\hat{\beta}_0$  is the intercept. The variables  $X_{1i}$  through  $X_{pi}$  represent the values of the factors for the  $i^{\text{th}}$  county and  $\hat{\beta}_1$  through  $\hat{\beta}_p$  are the coefficients that were estimated using the least squares regression method (19,21). To explore the data, 4 linear regression models were created, which differ by factors included in the model.

The first regression model included all 8 factors as prediction variables resulting in Model 1. The regression model was then fitted by stepwise variable selection in both the forward and backward direction by using AIC as a model selection criterion. AIC rewards goodness of fit and penalizes the model's complexity (19). This was done by using the R package MASS version 7.3–54 (R Foundation for Statistical Computing) (22). The simplified model resulted in Model 2. At this point, a decision was made to remove the incidence rate from the data to better see how the other sociodemographic factors contributed to breast cancer outcomes, resulting in Model 3. Model 3 was then fitted with stepwise variable selection, yielding Model 4. To best compare the expected mortality to the observed mortality with all predictors available, Model 2 was chosen as the final model. Model 2 was then used to predict the expected mortality rates for the second standardized incidence ratio (SIR) that was computed ( $SIR_{LM}$ ).

### Standardized incidence ratio

A SIR was used to compare the spatial distribution of counties in terms of mortality rates due to breast cancer. In general, the SIR compares the expected value of deaths to the observed value of deaths in a county. This is calculated with  $SIR_j = \frac{O_j}{E_j}$  where  $O_j$  is the observed value for county  $j$ , and  $E_j$  is the expected value for county  $j$ . A SIR of greater than 1 means that there were more deaths than expected, whereas a SIR of less than 1 means that there were fewer deaths than expected for that county.

Because age-specific mortality rates by county were not available, the observed age-adjusted counts per county were computed from

the age-adjusted rates per county as follows:  $O_j^* = \frac{O_j}{100,000} y_j$  and  $E_j^* = \left( \frac{\sum_{j=1}^{66} O_j}{100,000} \cdot y_j \right) \cdot \frac{y_j}{\sum_{j=1}^{66} y_j}$ , where  $O_j$  is the  $j^{\text{th}}$  county's ob-

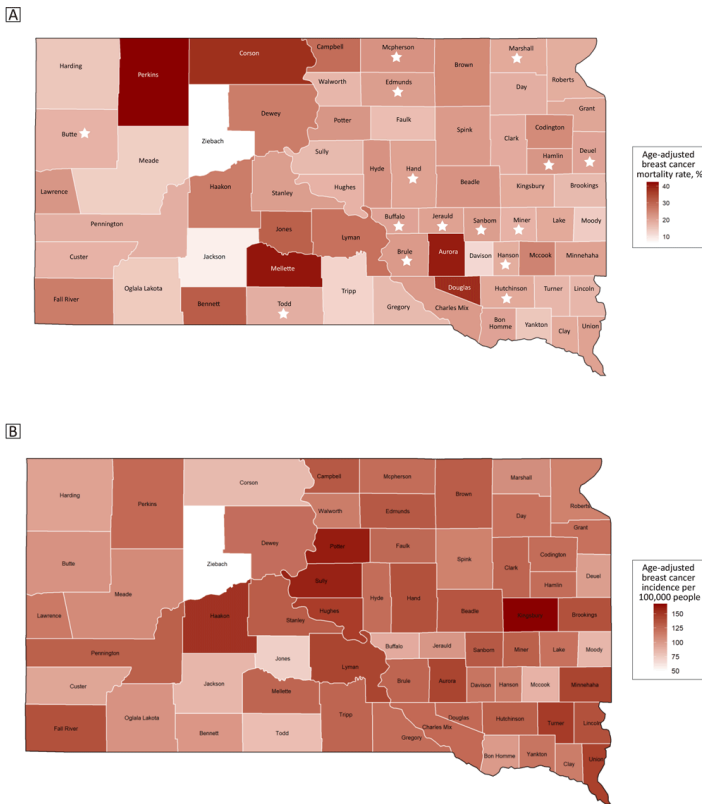
served age-adjusted mortality rate and  $y_j$  is the  $j^{\text{th}}$  county's population. The expected count was computed to be on the same scale as the observed count. These manipulations were used only to calculate an age-adjusted count SIR, referred to as  $SIR_{COUNT}$ . All other analysis of the data was completed with the original variable ( $O_j$ ), as described in the data source.

A SIR was calculated on the mortality in each South Dakota county ( $N = 66$ ). To account for the age adjustment of the data, 2 different SIRs were found: using the age-adjusted mortality count for South Dakota and using the expected rate obtained from the linear regression model. The first SIR accounted only for the age adjustment, and the second SIR accounted for more factors related to breast cancer. For example, the first SIR used the age-adjusted mortality count for a county for both observed and expected ( $SIR_{COUNT}$ ). The second SIR ( $SIR_{LM}$ ) used the mortality rates by county, which were per 100,000 persons and age-adjusted to the 2000 US standard population, and the South Dakota estimated population for observed and predicted mortality rate from the linear regression model for expected, which accounted for incidence rates and educational attainment. All statistical analyses used R version 4.1.2 and RStudio version 2021.09.2 build 382 (R Foundation for Statistical Computing) (23,24).

## Results

### Exploratory data analysis

To learn more about the data, we performed an Exploratory Data Analysis (EDA) using several techniques. We explored the geographic distribution of the breast cancer mortality rates by using a choropleth map of South Dakota with age-adjusted mortality and incidence rates (Figure 1). The eastern side of the state had lower and more consistent mortality rates, followed by the far western part of the state. The central west part of the state exhibited higher mortality rates.



**Figure 1.** Map A shows the age-adjusted breast cancer mortality rates and Map B shows the age-adjusted breast cancer incidence rates, by county (N = 66), South Dakota, 2001–2015. Counties whose mortality rates have been imputed are marked with a star. Source: South Dakota State Cancer Registry, South Dakota Department of Health (12).

Five counties had higher mortality rates than the rest of the counties: Perkins, Mellette, Aurora, Douglas, and Corson. Two counties had lower mortality rates than the rest of the counties: Ziebach and Jackson. Eight counties (Bennet, Buffalo, Corson, Jackson, Mellette, Oglala, Todd, and Ziebach) had a poverty percentage greater than 30%, which was distinctly higher than the others; each county contains land on an American Indian Reservation. Minnehaha and Pennington counties had high screening rates; these counties are home to the first- and second-largest cities in South Dakota, respectively, so they also had the greatest number of screening providers.

All variables, besides the number of providers and screening rate, had a correlation greater than zero with mortality rate. Incidence, educational attainment, and uninsured percentage all had a low positive correlation with mortality rate. The highest correlation with mortality rate was education attainment, at a value of 0.26. The remaining variables had a negligible correlation with mortality rate.

## Regression analysis

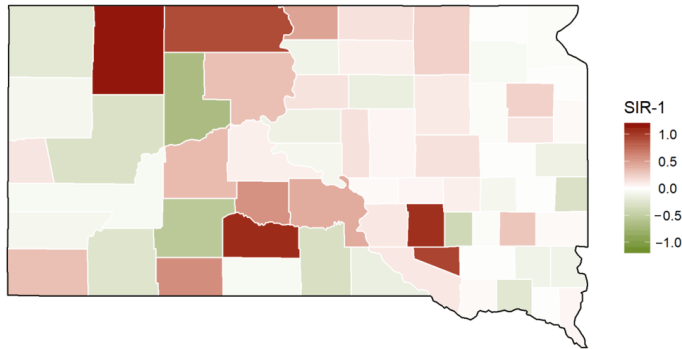
Model 2 had the highest adjusted  $R^2$  value (0.10) and the lowest AIC (441.79), with 2 significant factors associated with mortality rate; Model 3 had the lowest adjusted  $R^2$  (.004) and the highest AIC (453.32), with no significant predictors of mortality. In Model 2, breast cancer incidence and educational attainment were predictors of breast cancer mortality, indicating that as more people are diagnosed with breast cancer and as the percentage of people with less than a bachelor’s degree increases, breast cancer mortality rate increases (Table 2). Educational attainment was a predictor of mortality in all models, and the educational attainment statistic had a  $P < .001$ .

## Standardized incidence ratio

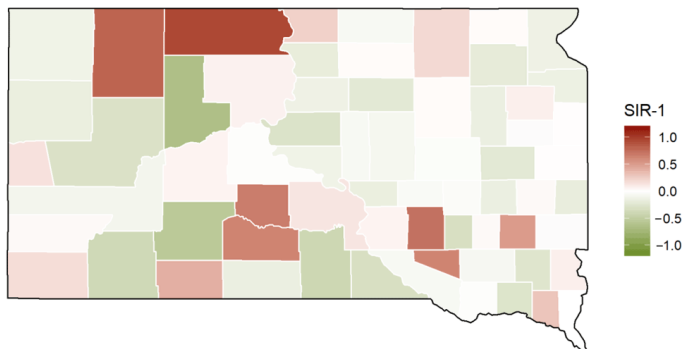
Thirty-five of the 66 counties had a  $SIR_{COUNT}$  greater than 1, meaning that more than half of the counties had more deaths than expected (Figure 2). The 5 counties with the highest  $SIR_{COUNT}$ , in decreasing order, were Perkins, Mellette, Aurora, Douglas, and Corson. Of those, Perkins, Mellette, and Aurora had more than twice the expected number of deaths. Ziebach, Jackson, Davison, Tripp, and Meade counties had the lowest  $SIR_{COUNT}$ , with Ziebach and Jackson both being less than half the expected number of deaths. The highest  $SIR_{COUNT}$  was 2.15 and the lowest was 0.31.

The opinions expressed by authors contributing to this journal do not necessarily reflect the opinions of the U.S. Department of Health and Human Services, the Public Health Service, the Centers for Disease Control and Prevention, or the authors’ affiliated institutions.

A Count SIR-1



B LM SIR-1



**Figure 2.** Map A shows the predicted breast cancer mortality rate of South Dakota counties, accounting for age-adjustment of the data, and Map B shows the predicted breast cancer mortality rate of South Dakota counties, accounting for age-adjustment, incidence rate, and educational attainment. Abbreviation: SIR, standardized incidence ratio. Map Sources: South Dakota State Cancer Registry, South Dakota Department of Health (12), Holzhauser et al (13), and the US Census Bureau (8).

On the other hand, for the  $SIR_{LM}$ , most counties had fewer deaths than expected, with 28 of 66 counties having a  $SIR_{LM}$  greater than 1. The 5 counties with the highest  $SIR_{LM}$  were similar to the  $SIR_{COUNT}$ : Corson, Perkins, Aurora, Jones, and Mellette in decreasing order. No county had more than twice the expected number of deaths. The 5 counties with the lowest  $SIR_{LM}$  were Ziebach, Jackson, Tripp, Oglala Lakota, and Davison. Ziebach and Jackson counties had less than half the number of expected deaths. The highest  $SIR_{LM}$  was 1.94 and the lowest  $SIR_{LM}$  was 0.35.

The results of the SIRs ( $SIR_{COUNT}$  and  $SIR_{LM}$ ) are presented in Figure 2. The eastern side of the state showed similar SIR values while the western side of the state had more variation. Perkins, Mellette, Aurora, Douglas, and Corson counties had  $SIR_{COUNT}$  values that were much higher than those of the rest of the counties. Ziebach and Jackson counties had the lowest  $SIR_{COUNT}$  values.

These counties with the highest  $SIR_{COUNT}$  made up 5 of the 6 counties with the highest  $SIR_{LM}$  values, with Jones County replacing Douglas County. The 2 counties with the lowest  $SIR_{COUNT}$  values were the same 2 counties with the lowest  $SIR_{LM}$  values.

## Discussion

Overall, we found a significant association between incidence rate and educational level with respect to breast cancer mortality rates. Breast cancer incidence was positively associated with mortality rates in South Dakota, which suggests that more breast cancer cases are associated with more breast cancer deaths. In addition, educational attainment was repeatedly identified as a significant factor for mortality. Gadeyne et al found inconclusive results in their study of breast cancer mortality and education; however, Albano et al found a significant association between educational levels and cancer in general (7,25), specifically that lower educational attainment was related to higher cancer mortality rates, reflecting the findings of this study. Race, median age, and number of women screened were not selected in the feature selection during stepwise regression in our study; similarly, race, median age, and number of women screened were not significant in our full model.

The 2020–2021 South Dakota Department of Education yearly review stated that American Indians were the largest minority group in school. However, American Indians still have a 63% completion rate for high school graduation and 59% attendance rate, compared with Whites who have a 94% completion rate for high school graduation and 94% attendance rate (26). An interesting point to consider is that South Dakota has no set standards for sex education (27). Thus, students are not taught reproductive health in general, including the importance of breast examinations, Pap smears, or prostate examinations. We advocate that set and scientifically backed health standards in high school would expose students at an early age to the risks of breast cancer and their options for screening.

The western half of South Dakota had more variability in SIR values, and the state's demographics could be a possible explanation. The 4 counties with the highest SIRs for both count and linear model SIRs were Corson, Perkins, Aurora, and Mellette, which are either in an American Indian reservation or neighbor a county within an American Indian reservation. Research on 3 tribes in western South Dakota supported that trust is often a barrier for American Indians (as are remote location and approvals by Indian Health Service programs) (28). Research in New Mexico reported that even after in-depth implementation of screening programs that lowered the barriers of cost, availability, and access to Native American and Hispanic women, the screening rates remained low,

under 40% of women annually (26). The high SIRs in or neighboring reservation counties may mean that trust is also an issue, and South Dakota has more to work on than accessibility to Native Americans.

After controlling for incidence rate and educational attainment, the  $SIR_{LM}$  values became less variable. The SIRs' decrease in range and mean closer to 1 indicate that the factors did affect mortality rate. This again agrees with findings from Albano et al that educational attainment affects mortality (7).

We found that some counties had a higher mortality rate than expected based on the age of the women in the county. Ziebach and Jackson counties had the highest mortality rates, and the counties with the lowest SIRs are not home to major medical centers. Haakon County is vertically between Ziebach and Jackson counties and is one of the counties that does not have a provider; however, Haakon County has a higher mortality rate than the average of counties of South Dakota and both SIRs greater than 1, which means there were more deaths than expected. The areas of the map where there are dark green counties next to dark red counties are either on an American Indian reservation or neighbor an American Indian reservation. The differences between counties do not come from any singular cause, but rather due to variations in race, poverty levels, and population size.

Our study has limitations, primarily in the absence of portions of mortality data. Because South Dakota is largely a rural state, several counties have small populations and see very few deaths from breast cancer. These numbers are then withheld from the public to protect the privacy of the patients. This suppression resulted in having to impute the mortality rates of 15 counties, possibly introducing errors. The assumptions of the regression model did not account for this error, which may confer bias on the results. The counties with American Indian reservations have another health system that could have resulted in the under-representation or over-representation of breast cancer deaths in those counties (29). In addition, a study found that the misclassification of the race of Native Americans caused an underestimation of mortality rates as well (30). These gaps in databases and their contents highlight research challenges that rural communities will continue to face when few data are collected, populations and incidence are sparse, and data are inconsistently collected by multiple sources. Sociodemographic data are also challenging to consistently collect throughout a state. More detailed data per county would help yield accurate and unbiased results. For example, considering education, Zajacova and Lawrence argue that education is not a single-generation factor (31). Having data on the educational attainment of a patient's parents or family, in addition to their own education-

al attainment, would allow us to assess the risk and see the relationship between education and breast cancer incidence or mortality. Thus, more research is needed to understand the effects education level has on financial security, stable employment, social success, and in turn, breast cancer mortality.

In conclusion, understanding the risk factors and geographic distribution of breast cancer mortality among women across the state will assist stakeholders with efforts at prevention and resource allocation guided by data.

## Acknowledgments

This research was partially supported by the Joseph F. Nelson mentorship and undergraduate award and a Van D. and Barbara B. Fishback honors college undergraduate Research Award from South Dakota State University. The contents are solely the responsibility of the authors. No copyrighted figures, surveys, instruments, or tools were used in this study.

## Author Information

Corresponding Author: Marisa Schulz, MSc, Mathematics and Statistics Department, South Dakota State University, AME Building, Box 2225, Brookings, South Dakota 57006. Telephone: 605-688-6316. Email: Marisa.schulz@jacks.sdstate.edu.

Author Affiliations: <sup>1</sup>Mathematics and Statistics Department, South Dakota State University, Brookings, South Dakota. <sup>2</sup>Sanford Research, Research Design and Biostatistics Core, Sioux Falls, South Dakota.

## References

1. American Cancer Society. Cancer facts and statistics. Accessed March 21, 2022. <http://cancerstatisticscenter.cancer.org/>
2. Centers for Disease Control and Prevention. United States Cancer Statistics (USCS) data visualizations. Accessed September 13, 2022. [https://gis.cdc.gov/Cancer/USCS/?CDC\\_AA\\_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcancer%2Fdataviz%2Findex.htm#AtAGlance/](https://gis.cdc.gov/Cancer/USCS/?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcancer%2Fdataviz%2Findex.htm#AtAGlance/)
3. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA Cancer J Clin* 2017;67(6):439–48.
4. Armstrong K, Eisen A, Weber B. Assessing the risk of breast cancer. *N Engl J Med* 2000;342(8):564–71.
5. DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, et al. Breast cancer statistics, 2019. *CA Cancer J Clin* 2019;69(6):438–51.

6. Hussain SK, Altieri A, Sundquist J, Hemminki K. Influence of education level on breast cancer risk and survival in Sweden between 1990 and 2004. *Int J Cancer* 2008;122(1):165–9.
7. Albano JD, Ward E, Jemal A, Anderson R, Cokkinides VE, Murray T, et al. Cancer mortality in the United States by education level and race. *J Natl Cancer Inst* 2007;99(18):1384–94.
8. US Census Bureau. Office of Management and Budget in statistical policy directive 14. Accessed August 2, 2022. [https://www.census.gov/topics/income-poverty/poverty/about/history-of-the-poverty-measure/omb-stat-policy-14.html#:~:text=RelatedSites-,OfficeofManagementandBudget\(OMB\)inStatisticalPolicyDirective,andestablishmentsforstatisticalpurposes](https://www.census.gov/topics/income-poverty/poverty/about/history-of-the-poverty-measure/omb-stat-policy-14.html#:~:text=RelatedSites-,OfficeofManagementandBudget(OMB)inStatisticalPolicyDirective,andestablishmentsforstatisticalpurposes)
9. Olson J, Cawthra T, Beyer K, Frazer D, Ignace L, Maurana C, et al. Community and research perspectives on cancer disparities in Wisconsin. *Prev Chronic Dis* 2020;17:200183.
10. South Dakota Department of Health. South Dakota comprehensive cancer control state plan, 2015–2020.
11. Centers for Disease Control and Prevention. National Breast and Cervical Cancer Early Detection Program screening program summaries: South Dakota. Accessed March 22, 2022. [https://www.cdc.gov/cancer/nbccedp/data/summaries/south\\_dakota.htm](https://www.cdc.gov/cancer/nbccedp/data/summaries/south_dakota.htm)
12. South Dakota cancer incidence and mortality. Accessed March 22, 2022. <https://www.sdancerstats.org/>
13. Holzhauser C, Da Rosa P, Michael S. Forecasting participants in the All Women Count! mammography program. *Prev Chronic Dis* 2018;15:180177.
14. South Dakota Department of Health. All Women Count! cancer control program. Accessed March 21, 2022. <https://getscreened.sd.gov/count/>
15. US Census Bureau. Small Area Health Insurance Estimates (SAHIE) program. Accessed August 2, 2022. <http://www.census.gov/programs-surveys/sahie.html>
16. Wasserman L. Classification. In: *All of statistics: a concise course in statistical inference*. New York (NY): Springer; 375 pages.
17. Peterson LE. K-nearest neighbor. *Scholarpedia* 2009;4(2):1883.
18. Kowarik A, Templ M. Imputation with the R Package VIM. *J Stat Softw* 2016;74(7):1–16.
19. Bruce PC, Bruce A, Gedeck P. *Practical statistics for data scientists: 50+ essential concepts using R and Python*. Second edition. Sebastopol (CA): O'Reilly Media, Inc; 2020.
20. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Parzen E, Tanabe K, Kitagawa G, editors. *Springer series in statistics. Second International Symposium on Information Theory*; 1973. Pages 267–281.
21. Kutner MH, Nachtsheim C, Neter J. *Applied linear regression models*. New York (NY): McGraw-Hill/Irwin; 2004.
22. Venables WN, Ripley BD, Venables WN. *Modern applied statistics with S*. 4th edition. New York (NY): Springer; 2002.
23. R: the R project for statistical computing. Accessed March 22, 2022. <https://www.r-project.org/>
24. RStudio. Open source and professional software for data science teams. Accessed March 22, 2022. <https://www.rstudio.com/>
25. Gadeyne S, Menvielle G, Kulhanova I, Bopp M, Deboosere P, Eikemo TA, et al. The turn of the gradient? Educational differences in breast cancer mortality in 18 European populations during the 2000s. *Int J Cancer* 2017;141(1):33–44.
26. Paskett ED, Tatum C, Rushing J, Michielutte R, Bell R, Foley KL, et al. Racial differences in knowledge, attitudes, and cancer screening practices among a triracial rural population. *Cancer* 2004;101(11):2650–9.
27. South Dakota Department of Education. South Dakota health education standards; 2018. Accessed July 24, 2022. <https://doe.sd.gov/board/packets/documents/112017/item5doc2.pdf>
28. Rogers D, Petereit DG. Cancer disparities research partnership in Lakota Country: clinical trials, patient services, and community education for the Oglala, Rosebud, and Cheyenne River Sioux tribes. *Am J Public Health* 2005;95(12):2129–32.
29. Office of Minority Health. Profile: American Indian/Alaska Native. Published 2018. Accessed March 25, 2022. <https://www.minorityhealth.hhs.gov/omh/browse.aspx?lvl=3&lvlid=62>
30. Espey DK, Jim MA, Richards TB, Begay C, Haverkamp D, Roberts D. Methods for improving the quality and completeness of mortality data for American Indians and Alaska Natives. *Am J Public Health* 2014;104(Suppl 3):S286–94.
31. Zajacova A, Lawrence EM. The relationship between education and health: reducing disparities through a contextual approach. *Annu Rev Public Health* 2018;39(1):273–89.

## Tables

**Table 1. Factors for Regression Analysis, Study on Breast Cancer Mortality in South Dakota, 2001–2015**

Factor	Description
Breast cancer incidence rates	Age-adjusted breast cancer incidence rate per 100,000 persons
Breast cancer screening rates	Estimated mammography screening rates
Number of screening providers	The number of medical providers per county that provide breast cancer screening
Poverty level	The percentage of residents living in poverty per county
Insurance status	The percentage of residents without any insurance
Median age	Median age of the county's residents
Race	The percentage of residents that are non-White
Educational attainment	The percentage of the population aged 25 years or older with less than a bachelor's degree

---

The opinions expressed by authors contributing to this journal do not necessarily reflect the opinions of the U.S. Department of Health and Human Services, the Public Health Service, the Centers for Disease Control and Prevention, or the authors' affiliated institutions.

---



**Table 2. Regression Models, Study on Breast Cancer Mortality in South Dakota, 2001–2015**

Coefficient	Model 1	Model 2	Model 3	Model 4
	Estimate (95% CI)			
Intercept	-15.375 (-39.388 to 8.638)	-11.113 (-32.253 to 10.026)	-2.248 (-23.425 to 18.929)	4.114 (-12.158 to 20.386)
Median age	-0.059 (-0.481 to 0.363)	–	-0.043 (-0.478 to 0.391)	–
Non-white percentage	-15.119 (-37.725 to 7.486)	–	-11.266 (-34.233 to 11.700)	–
Poverty percentage	0.208 (-0.292 to 0.708)	–	0.030 (-0.455 to 0.515)	–
Uninsured percentage	0.532 (-0.275 to 1.340)	–	0.418 (-0.406 to 1.242)	–
Number of providers	0.210 (-1.318 to 1.738)	–	0.458 (-1.097 to 2.012)	–
Screened	-0.002 (-0.013 to 0.009)	–	-0.003 (-0.015 to 0.008)	–
Incidence	0.101 <sup>a</sup> (0.006 to 0.197)	0.085 <sup>a</sup> (0.007 to 0.163)	–	–
Educational attainment	0.286 (-0.100 to 0.672)	0.333 <sup>b</sup> (0.088 to 0.578)	0.312 (-0.085 to 0.708)	0.258 <sup>a</sup> (0.016 to 0.499)
<b>AIC</b>	450.30	441.79	453.32	444.55
<b>Adjusted R<sup>2</sup></b>	0.06	0.10	0.004	0.05

Abbreviation: – , not applicable; AIC, Akaike information criterion.

<sup>a</sup> P = .01.

<sup>b</sup> P = .001.