

III. TECHNICAL NOTES

Age-adjusted incidence rate: Age-adjusted incidence rates are calculated using the direct method and standardized to the age distribution of the 2000 US standard population (Appendix A). Age adjustment allows rates for one geographic area to be compared with rates from other geographic areas that may have differences in age distributions. Any observed differences in age-adjusted incidence rates between populations are not due to different age structures. Reports prior to 1999 used the 1970 US standard population. In conformity with the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) Program guidelines, the incidence rates for cancer sites exclude the following:

- *In situ* cases, except bladder
- Basal and squamous cell skin cancers
- Cases with unknown age
- Cases with unknown gender

Age-adjusted mortality rates: Mortality rates are calculated for total cases and separately for males and females. The mortality rates are age-adjusted to the 2000 US standard population using five-year groups and are per 100,000 persons. Rates are presented for 2017 and for the five-year period, 2013-2017.

Age-specific incidence rates: Age specific rates are calculated by dividing the number of cases for a given age group by the total population of that age group and are expressed as an average annual rate per 100,000 persons by age group. Age specific rates exclude the same types of cases that are excluded from age-adjusted incidence rates. These rates, however, are crude rates, i.e. not age-adjusted.

Annual percent change (APC): The annual percent change is the average rate of change in a cancer rate per year in a given time frame indicating how fast or how slowly a cancer rate has increased or decreased each year over a period of years. A negative APC describes a decreasing trend, and a positive APC describes an increasing trend.

Average years of life lost (AYLL): This is the extent to which life is cut short due to premature death. This is obtained by dividing the years of potential life lost (YPLL) by the number of deaths. On average each person who dies from cancer loses 15 years of their life.

Cancer case definitions: A "cancer case" is defined as the primary cancer site, i.e., the site where the cancer started. Since an individual can have more than one primary cancer site, the number of incident cancer cases could be greater than the number of persons who are diagnosed with cancer. A *metastasis* is not a primary site.

Changes in diagnostic criteria: Early detection resulting from either screening or early response to symptoms may result in increasing diagnosis of small tumors that are not yet life-threatening. This may raise incidence and survival rates but without changes in mortality rates. Cancers likely to be affected are breast, colon, cervix uteri, prostate, and melanoma. Prostate cancer is particularly prone to changing diagnostic criteria.

Confidence intervals (CI): A confidence interval tells how confident we are of the accuracy of the calculated rates. The SDCR uses a computed interval with a given probability of 95%, i.e., the true value of the calculated rate is contained within the interval. Thus, given a calculated rate of 191.4 and a confidence interval of 182.1 to 200.8, it is better to say that the true rate will fall between 182.1 and 200.8. The larger the sample size, the shorter the interval size, giving us more certainty that the rate is correct. When CI for percentages contains zero, the rate is considered to be stable. Above zero, the statistical significance is higher and below zero it is lower.

Data source: All data, tables, and figures come from the South Dakota Department of Health, *American Cancer Society Facts and Figures 2017* or *SEER Cancer Statistics Review 1975-2016* and should be cited as such if taken out of this report in part. SEER data represents approximately 10% of the U.S population.

Disparity: Health disparities are differences in the incidence, prevalence, mortality, and burden of diseases and other adverse health conditions that exist among specific population groups in the United States.¹ Health disparities can be defined as a specific group bearing a disproportionate share of negative health outcomes compared to the general population, i.e., disease, disability, and death.² Disparity can occur as a result of factors such as poverty, living in geographically underserved areas and belonging to specific minority groups.

Early detection/screening: Improved early detection/screening may produce increases in both incidence and survival rates. Increases may occur as a result of the introduction of new procedures. The interval between the time a cancer is diagnosed by a screening procedure and the time when it would have been diagnosed in the absence of screening procedures is called the lead-time. Changes in lead-time, for example, in breast cancer diagnosis, have led to increased survival rates and reduction of mortality.

Limitations to data interpretation and comparison: A number of factors need to be considered when reviewing cancer statistics and interpreting them. A cancer registry database is a fluid and dynamic database, therefore, the reported number of new cases in a particular race, gender, and age cancer category may change for the calendar year for which the data have already been reported in a previous publication. Additional cancer cases which have been previously overlooked for a given diagnosis year may be found and reported to the central registry. There may also be elimination of duplicate records for the same patient, often due to name changes or spelling corrections.

Metastasis: When cancer spreads from the primary site to other organs or tissues of the body, it is said to metastasize. Cancer usually spreads through the blood or the lymphatic system.

Mortality/incidence ratio (M/I): This ratio is calculated by dividing the number of deaths in a given year by the number of new cancers diagnosed in the same year. The death to case ratio provides a crude indication of the prognosis for patients. A ratio approaching 1.0, when the number of deaths equals the number of cases for a particular type of cancer, indicates a poor prognosis. A lower ratio indicates fewer deaths relative to the number of cases and suggests a better prognosis.

Percent change: This is the difference between two rates expressed as a percentage.

Racial misclassifications: When race is not specified in a source document and the default is to record these cases as white or unknown, the results are considered biased. Numerator error can occur because of misclassification.

Rate comparisons: When comparing age-adjusted rates and age-specific rates based on fewer than 10 cases, rate comparisons are difficult to interpret. In comparing rates among geographic areas such as counties, states and health districts, the absolute numbers and differences in demographics should be considered, as well as clinical significance of the disease. Data quality indicators for each registry should also be reviewed. Interpretations made without considering these factors may be misleading. There will also be differences between mortality statistics published by various agencies and the mortality rates in this report.

Risks and associated risk factors: These were developed using the "American Cancer Society Textbook of Oncology," and the Harvard Cancer Center, Causes of Human Cancer.

Stage at time of diagnosis: Staging is the process of describing the extent or spread of disease from the origin, which is the primary site. Summary staging is the standard used for comparison nationally. SEER Summary Stages 2000 are defined as follows:

- **In Situ:** Malignant cells are within the cell group from which they arose, without penetration of the basement membrane of the tissue and no stromal invasion. *In situ* is “in place”.
- **Localized:** The malignant cells are limited to the organ of origin and have spread no farther than the organ in which they started.
- **Regional:** The tumor is beyond the limits of the organ of origin by direct extension to adjacent areas with or without lymph node involvement.
- **Distant:** The primary tumor has broken away and has traveled, growing secondary tumors in other parts of the body. It has metastasized.

In situ and localized stages are the **early stages** of diagnosis. Regional and distant stages are **late stage** diagnoses.

Staging: Stage is based on an assessment of the size of the primary tumor, whether it has spread, and, if so, how far. Because an accurate diagnosis is so important to effective treatment, physicians might use physical exams, imaging, lab tests, a biopsy, an analysis of the patient’s body fluids, and surgery in various combinations in the staging process. Advancement in diagnostic procedures may change in due time. These advancements might increase the chance that a given cancer will be diagnosed at a more advanced stage, for example with new scanning methods metastases can be detected. Therefore, if someone was previously diagnosed with a localized tumor, they may now be staged as distant. This is called stage migration and can affect the analysis of all solid tumors.

Statistical significance: This determines whether an event happens by chance alone. The null hypothesis states that in a given place and a period of time, all events occur randomly by chance. If not, then there is statistical significance. Confidence intervals are used to test statistical significance in this report. If the confidence intervals of two different rates intersect each other, then there is no statistical difference between the two rates.³ However, if the confidence intervals do not intersect one another, there is statistical significance. This report looks at the South Dakota rates as compared to the US national rates using SEER data.

In South Dakota, case counts can be very low; therefore, magnitude bias is inherent with confidence intervals and z- tests. For example, in the year 2001, cervical cancer rates were 10 per 100,000 American Indian women, a cervical cancer age-adjusted rate six times higher than white women in South Dakota. However, the case counts were two for American Indians and 10 whites. Small numbers result in wider confidence intervals, thus less confidence in the data.

Years of potential life lost (YPLL): The years of potential life lost is calculated for each individual who dies of a cancer of interest by determining the number of years of additional expected life if that person had lived to 75 years. The YPLL in the general population associated with a particular cancer is the sum of this expectation over all those individuals who died of that cancer in a particular year. YPLL reflects the burden of cancer on younger persons while mortality rates reflect the burden on older persons.

¹<http://epi.grants.cancer.gov/ResPort/HDOverview.html>

²<http://www.omni.org/docs/CMHFPProceedings.pdf>

³BIostatistics *The Bare Essentials*, 2nd edition Norman and Shreiner Page 512