IX. Technical Section

A. Data Sources

The Oregon data in this annual report come from a variety of sources:

Oregon Incidence Data

All cancer incidence data were obtained directly from the Oregon State Cancer Registry. There have been a number of changes in cancer incidence reporting and collection that were implemented in 2001 including transitions to Summary Stage 2000 and ICD-O-3. These changes create problems when comparing data across years. However, the updated reporting requirements and coding guidelines are intended to reflect current medical knowledge of the behavior, pathology, prognosis, treatment of cancers, and, hence, increase the applicability of registry data for surveillance and research.

Specifically, ovarian cancer, lymphomas, leukemias, and other hematopoietic disease incidence will be difficult to compare across years due to changes in reportability or definition beginning in 2001 based on new ICD-O-3 definitions. Also, comparing stage data for lung, ovarian, and colorectal cancers across the years is problematic based on the new Summary Stage 2000 guidelines.

Oregon Mortality Data

Cancer mortality data include all cancer deaths among Oregon residents. These data exclude cases in which benign or *in situ* neoplasms or those of unknown behavior are listed as the cause of death with the exception of the newly reportable cases for 2001. See the *What's New in 2001?* section for additional information.

All of the cancer mortality data were obtained from the Center for Health Statistics and Vital Records (CHS) death certificate data. CHS is the state's depository for all vital records and is a major information source for vital statistics and health survey data about Oregonians. Mortality data from this report is age-adjusted and is, therefore, not comparable to the crude rates published by CHS. Also, invalid age or cause of death codes are excluded, so mortality counts presented in this report may differ from data presented in CHS publications.

There has been a methodological change that alters how mortality data are presented but does not represent a true variation in the underlying burden of cancer in Oregon. In

1999, all state vital records departments were required to switch from the ninth revision of ICD to the tenth for recording cause of death. The *International Classification of Diseases (ICD)* is the classification used to code and classify mortality data from death certificates. The ICD has been revised about every ten years since its inception in 1900. The intent of the revisions is to better reflect medical advances in disease nomenclature and etiology. ICD-10 is far more detailed than ICD-9 with approximately 3,000 additional codes. ICD-10 also uses an alphanumeric system, whereas ICD-9 is numeric only. The ICD-10 system is compatible with the ICD-Oncology (ICD-O) system used for reporting cancer cases, based on topography, whereas the ICD-9 system was not completely compatible.

Although the ICD-10 and ICD-O systems correspond more directly with each other, switching to the ICD-10 system creates issues that are crucial to the interpretation of the mortality data over time. Specifically for cancer deaths, the number of overall cancer deaths using the ICD-9 and the ICD-10 systems is essentially the same. However, causes of death are moved among different categories.

The most notable change for cancer mortality coding involves lung cancer deaths. Lung cancer is classified as secondary to some other cancers by ICD-10. Therefore, some of the lung cancer deaths classified by ICD-9 were moved to 1 of 15 other primary sites by ICD-10. The result is a lower number of cancer deaths defined as lung cancer. *This is an artifact due to a change in classification and not a true change in mortality.*

The second most noticeable change for cancer death coding involves cancers with multiple primary sites. Using the ICD-9 system, individuals who died with two primary cancer sites were classified based on the site first listed on the death certificate. Under the new ICD-10 system, the same individuals are coded into the miscellaneous site group. The result is a higher rate of cancer death determined as miscellaneous cancer site and lower rates of cancer in sites with common secondary cancers such as oral cancer. *This is an artifact due to a change in classification and not a true change in mortality rates.*

The third most observable change in cancer coding using ICD-10 concerns mesothelioma. Mesothelioma is a rare form of cancer with cancerous cells found in the lining of the chest (pleura) or abdomen (peritoneum). This cancer was indistinguishable from respiratory and digestive system cancers with the ICD-9 system. The new ICD-10 coding creates a separate site code for these cases. This affects primarily the number of respiratory system cases, which comprise the majority of malignant mesotheliomas.

Oregon Screening Data

The cancer screening data were obtained from the Behavior Risk Factor Surveillance System (BRFSS) maintained by CHS. The BRFSS is an annual telephone survey of adults concerning health-related behaviors. The information collected is used to guide health promotion and disease prevention programs. The BRFSS includes questions on health behavior risk factors such as seat belt use, diet, weight control, tobacco and alcohol use, physical exercise, preventive health screening, and use of preventive and other health care services.

Oregon Population Data

Denominators used to calculate Oregon incidence and mortality rates are population estimates from the Center for Population Research and Urban Studies, Portland State University (PSU), a resource center for population data designated by the US Bureau of the Census.

In 2000, we had a change in denominator data with the availability of updated US Census data. Population data based on the 1990 US Census were used for 1996-1999 rate calculations and population data based on the 2000 US Census were used for 2000-2001 rate calculations. There is a dramatic increase in population from 1999-2000 using 1990 estimates. This results in an artificial deflation of the 2000 rates when compared to earlier years.

In the future, intercensal estimates for years 1996-1999 will be used, which more accurately describe the true Oregon population. However, at the time of publication, intercensal estimates at the level of detail necessary for calculating age-adjusted rates (18 age groups at county level) were not available.

Oregon cancer rates by race and ethnicity are calculated using US Census data for race population denominators. Because the estimates differ slightly from the estimates of total population from PSU, these rates cannot be used for comparison with the *All Races, All Ages Rates*. The estimates are based on the US 2000 Census modified age, race, sex (MARS) file that contains multiple race categories. The MARS file was then bridged, using allotment estimates released by the National Center of Health Statistics. (*See Appendix E.*) The distribution was based on California studies documenting which single race a mother chose for her multiracial child. The race group denominators were then estimated for years 1996-1999 and 2001 based on total change from 2000 census data.

National Data

National incidence data were calculated from the Surveillance, Epidemiology and End Results (SEER) Cancer Incidence Public-Use Database, 1973-1999, Department of Health and Human Services, Public Health Service, National Cancer Institute. The SEER data were used to calculate a five-year aggregate rate for 1995-1999.

Additional national incidence data were obtained from *US Cancer Statistics Working Group, United States Cancer Statistics: 1999* Incidence, Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute: 2002.

National mortality data were calculated using the CDC Wonder website at: http://wonder.cdc.gov/. National trend comparison data were obtained from Weir, HK et al. Annual Report to the Nation on the Status of Cancer, 1975-2000, Featuring The Uses of Surveillance Data for Cancer Prevention and Control. *Cancer*. 2003 Sep 3;795(17):1276-1299.

National rates used in the Public-Use Database and CDC Wonder were calculated using national census 2000 population data for denominators.

B. Data Quality

Data Review

When OSCaR receives reports, they are closely reviewed and edited for quality control. The first cases reported contained many inconsistencies, as hospital abstractors were still becoming familiar with OSCaR reporting requirements. Currently, these systematic problems have been largely resolved.

The anatomical site of the origin of the cancer was known and reported for 97.7% of reported cancers in 2001. The stage of progression of the cancer at the time of diagnosis was determined for 89.9% of reported cancers in 2001.

The accuracy and usability of OSCaR data is constantly increasing through efforts on many different levels. Data review protocols to review Hispanic ethnicity for American Indians and Filipinos have been instituted to increase accuracy of ethnicity data. A protocol for reviewing sex coding is currently being developed. Increased use of registry data in linkage projects through academic research as well as general registry operations like death clearance (see page 87) helps to ensure that Registry data are reviewed and corrected on many levels in addition to the standard internal data review protocols.

One notable effort is in race/ethnicity coding for American Indians/Alaskan Natives (AI/AN). Due to a cooperative effort among state registries, the Northwest Portland Area Indian Health Board, the Indian Health Service (IHS), and the Centers for Disease Control, OSCaR links with local tribal registry data and national IHS patient data to determined if AI/AN are miscoded as white or some other race. One-third of the AI/AN cases currently in the OSCaR database were identified as miscoded and have been corrected due to this effort. This linkage and correction occurs annually.

In addition to internal Registry operations, OSCaR also conducts audits of reporting hospital and facilities across the state to assess the quality and completeness of data maintained in the central cancer registry. Initially, hospitals to be audited were selected based on the identification of reporting problems (i.e., a high number of missed cases). However, because overall reporting has improved from early years, the Registry now performs random facility audits every quarter. The hospitals are divided into groups based on the total number of hospital patients and the sampling method used ensures that all sizes of facilities are selected each time for auditing. OSCaR is also beginning to audit other reporting facilities such as free-standing clinics and cancer centers.

C. Cancer Reporting

Reportable Cancers

Not all cancers diagnosed in Oregon are reportable to the Oregon State Cancer Registry (OSCaR). Cancers that are reportable include all malignant neoplasms that are *in situ* or invasive (ICD-O behavior codes 2 and 3) with the following exceptions:

- Basal and squamous cell carcinoma of the skin
- Carcinoma *in situ* of the cervix (CIN)

Although incidence is high for these cancers, they have an extremely high treatment success rate. In addition, *in situ* cervical cancer is a diagnosis that is inconsistently used. Carcinoma *in situ* of the cervix overlaps with some diagnoses that indicate precancerous conditions (i.e., High Grade SIL (Squamous Intraepithelial Lesion), CIN 3.

Reporting Source

By law, outside of the above exceptions, all cancers diagnosed or treated in Oregon must be reported to OSCaR by the patient's physician. However, most hospitals retain tumor registrars that are trained to collect and report cancer cases in accordance with national standards. Most of the cases included in this report were reported from hospitals by tumor registrars. However, some case reports originated from doctors' offices and a few from death certificates. Since cancer reporting started in 1996, 89.2% of new cancer diagnoses have come from hospitals, 8.1% from physician offices, and 1.8% were identified from review of death certificates. The remainder were identified by review of pathology reports from laboratories or by autopsy. Many of the physician office cases were initially identified through active follow-up from laboratory reports or death certificates, and the physician was queried for information.

Death Clearance

Death clearance is a registry process used to identify cases diagnosed by a physician but not reported to the Registry as well as cases diagnosed through autopsy. Death certificates are compared, or cleared, with registry files. Deaths due to cancer that are not found in the Registry are investigated further by contacting the certifying physician listed on the certificate. Cases with no physician response and subsequent report are classified as death certificate only (DCO) cases. Cases for which a response and full report are received are classified as a physician office report. Deaths due to cancer diagnosed prior to the Registry's first reporting year, 1996, are not included in the Registry.

Full death clearance procedures were not necessary during the first few years of Registry operation since most of the cancer deaths were due to cancers diagnosed prior to 1996. Initially, death clearance was performed only for selected cancer sites that have known short-term survival. Death clearance on all sites was first performed for 1999 deaths.

Initial death clearance efforts focused on deaths due to cancers of the esophagus, liver, lung, pancreas, stomach, multiple myeloma, and cancers with unknown primaries for years 1996-1998. For the year 1998, lung cancer deaths were aggressively reviewed in this process. For the year 1999, death certificate review procedures were expanded to include all cancer sites. Consequently, there was a higher percentage of cases reported from death certificates since 1999. Typically, cancer cases identified by death certificate are the cases with the worst prognosis with the stage being distant or unknown.

Due to increased review, more DCO cases were identified from 1999 to present. These cases differ from other cases due to the increased severity of disease and lack of information about the cases. DCO cases are staged unknown due to lack of information. This requires caution when comparing stage of diagnosis from year to year.

Changes in the death clearance process resulted in increased data quality and completeness for diagnosis year 2000 onward. Prior to 2000, Registry cases and the death certificate data were linked on Social Security Number in an external system. Any cancer deaths not reported in the Registry were followed back to the physician or hospital to see if they were missed cases. The death clearance process is now performed using an automated program within the main Registry system. The deterministic match criteria are extremely stringent and result in the identification of thousands of discrepancies in demographic data between death certificate data and the cancer registry. The discrepancies are reviewed and corrected.

This process improves completeness of race data reporting by supplying race information from the death certificate on cases in the Registry that were reported as *unknown* race. Also, sex data reporting improved. Discrepancies between the death certificate and registry data for sex of male breast cases prompted further review of male breast cancer cases. Twenty-four male breast cancer cases were reported in 1999. After review of each case, only 17 were actually male. Seven of the cases had been miscoded and were actually female.

Primary Site Definitions

Cancer data presented in the current incidence and mortality sections of this report follow nationally accepted standards for reporting such data. Cancer data are classified using SEER Recodes. (Please see *Appendices F and G*.)

For mortality in years 1996-1998, the ICD-9 codes did not directly match ICD-O codes. Therefore, minor discrepancies exist for those years between Vital Statistics counts and the mortality counts reported in this publication. Beginning in 1999, with the change to ICD-10 coding, the mortality coding and counts match exactly for most sites. However, in 2001, the all-site mortality and the miscellaneous cancer mortality differ due to the inclusion of the newly reportable cancers which are excluded from the Vital Statistics counts.

Timeliness

By law, all newly diagnosed reportable cancers must be reported to OSCaR by the hospital or by the patient's physician within six months after diagnosis. This allows information about the first course of treatment to be included in the Registry. Historically, OSCaR's database infrastructure did not allow tracking of the timeliness of reporting. However, a reprogramming of the central database now allows such tracking. Timeliness will be reported in subsequent annual reports.

Multiple Primaries

The majority of cancer diagnoses reported to OSCaR were the first primary cancer diagnosed for the patient. However, nearly 20% of the cancer diagnoses occur in individuals with a previous cancer, so the number of cancer cases and the number of people with cancer are not the same. Rates are calculated using cancer cases as the numerator and the population as the denominator, not the number of individuals with cancer.

Case Ascertainment/Completeness

The Registry staff conducts random casefinding audits to monitor case reporting completeness from hospitals and contacts physician offices that have a reduction in case reporting. Identifying missed cases through review of pathology reports and death certificates is part of normal Registry procedures. Data sharing agreements among neighboring states help identify Oregon patients diagnosed elsewhere.

The 2001 data have a greater than expected number of cases based on national models. However, using mathematical models based on national numbers to estimate completeness for Oregon has inherent limitations. The estimated percentage of case completeness is calculated by dividing the observed age-adjusted cancer incidence rate by the expected age-adjusted cancer incidence rate. The expected cancer incidence rate was generated by multiplying the 1996-2000 national cancer incidence-to-mortality rate ratio from the SEER Program by the Oregon age-adjusted cancer mortality rate for the respective year. One compatibility problem with local data is the use of national population figures as Oregon estimates instead of local population figures. Generally, the national estimates are lower than local estimates, which results in a lower number of expected cases using this model.

D. Epidemiologic Measures

Cancer Rates

In analyzing Oregon's cancer data, we looked at various measures commonly used in epidemiologic studies of cancer. One measure is a rate. Rates help compare the burden of disease across populations of various sizes.

The incidence rate provides information on the frequency with which cancers are occurring in the population. Incidence rates include cases of invasive cancer only in the rate calculation except in the case of cancer of the bladder, which includes *in situ* cases in the incidence rate calculations. The mortality rate describes the frequency of deaths due to cancer.

All rates in this report are per 100,000 population. Rates based on counts < 11 are likely unstable and are not presented.

Crude Rates – Crude rates are desirable when a summary measurement is needed and there is no need to adjust for confounding factors, such as age. Since cancer risk is very dependent upon age, age-adjusted rates are more useful measurements for comparison among regions, time periods, etc. Crude rates are not included in the tables in the annual report but are still reported for individual sites in the *Selected Sites* sections. The denominators in Figure 79 are used to calculate the crude rates.

Figure 79

Oregon's Population by Year			
Year	**Total**	**Male**	**Female**
1996	3,181,005	1,569,124	1,611,881
1997	3,217,003	1,587,870	1,629,133
1998	3,267,542	1,613,653	1,653,889
1999	3,300,802	1,629,899	1,670,903
2000	3,436,770	1,703,676	1,733,094
2001	3,471,704	1,721,068	1,750,636

Age-Adjusted Rates – Age-adjusted rates are calculated to allow comparisons between two different populations (i.e., Oregon and the US) whose age distributions differ. Age-adjusted rates are calculated by the direct method, using the age distribution of the 2000 United States Standard Population. All age-adjusted rates are expressed as events per 100,000 individuals per year.

In the past, a number of different standard populations have been used. Most vital statistics data were age-adjusted using a standard population based on the 1940 United States population. Most cancer data were age-adjusted using a standard population based on the 1970 United States population. Age-adjusted rates calculated using different standard populations are not comparable.

The Year 2000 Standard has a higher percentage of individuals in the middle and older age groups. Hence, more weight is applied to cancer cases or deaths of individuals in these age groups using the direct standardization method. These are the same groups that have higher numbers of cancer. Therefore, using the Year 2000 Standard for age-adjusting results in cancer rates that are higher than rates using outdated population standards.

Also, switching to a new population standard means reported rates are not comparable to rates in reports prior to 1999. All age-adjusted rates presented in this report are calculated using the Year 2000 Standard to ensure comparability.

Childhood Cancer Classifications

The classification system used here to record the occurrence of childhood cancer was developed by the World Health Organization's International Agency for Research on Cancer (IARC). This system places a greater emphasis on tumor morphology than does the International Classification of Diseases (ICD) classification system, which emphasizes tumor location.

Childhood cancers are defined as cancers diagnosed in individuals less than 15 years of age. The five-year age group stratification for childhood cancers is 0-4, 5-9, and 10-14. The following are included in the IARC classification system for lymphomas: Hodgkin lymphoma, non-Hodgkin lymphoma, Burkitt lymphoma, Histiocytosis X, unspecified lymphomas, and other reticuloendothelial neoplasms. Histologic confirmation is obtained on almost all diagnoses of childhood cancer reported to OSCaR.

Geographic Comparisons

County Comparisons – This report compares incidence and mortality rates across county geographic boundaries. These analyses may help target screening and/or educational efforts. Because some counties with small populations only have a few cases reported, rates for those counties are unstable and must be interpreted with caution.

Regional Comparisons – Regional maps depict "smoothed" or fitted county rates and should not be used to evaluate individual county rates. Data smoothing is a statistical technique intended to limit the influence of randomness in data. Through this process, information (cancer rates) is interpolated or "borrowed" from neighboring areas to stabilize results for less populated areas.

The statistical algorithm used for the regional maps is a weighted, median-based method intended for non-point, spatial data called head-banging. The observed rate for each county is compared to the median rates of neighboring counties. The county rates are weighted by population size to ensure that statistically stable rates are not modified based on rates from sparsely populated counties. Please see http://srab.cancer.gov/headbang/ for additional information.

This process stabilizes the county cancer rates for counties with low population density to allow potential, geographic patterns to emerge. However, the smoothed rates for individual counties are not appropriate for single county comparisons. To compare single county rates, please use the county rates presented in Tables 3 and 4.

Incidence Counts

All primary reportable malignancies diagnosed among Oregon residents are reported to OSCaR. Cases are categorized based on the International Classification of Diseases for Oncology (ICD-O) and are presented using the Surveillance, Epidemiology, and End Results (SEER) Program recodes. (*See Appendix F*.)

Cancer counts represent the number of primary cancers reported to the OSCaR, not the number of persons with cancers. People may be diagnosed with more than one primary tumor (e.g. lung cancer and Hodgkin Lymphoma), and, therefore, counted as more than one case. About 20% of the cases reported to OSCaR occur in a person who has already been diagnosed with cancer.

The *All Sexes* category used in this report for cancer incidence includes cases defined as male, female, and other (i.e., hermaphrodite, transsexual) and may exceed the total of male and female alone.

The number of cancers is reported in two ways – *total* cancers and *invasive* cancers. The invasive cancer category excludes *in situ* cancers with the exception of urinary bladder cancer. The total cancer category includes all cancers, regardless of stage at diagnosis, with the exception of cervical cancer since *in situ* cervical cancer is not reported to the Registry.

Prognosis and Burden Indicators

Several methods are used to measure the prognosis and burden of cancer within a defined population. In this report we use two such measures: the mortality to incidence ratio (M/I ratio) and the years of potential life lost (YPLL).

The M/I ratio provides a measure of disease severity. The M/I ratio is the number of deaths divided by the number of invasive cases (of that particular cancer). The closer a value is to 1.0, the poorer the prognosis for that cancer.

The years of potential life lost (YPLL) index quantifies premature mortality from cancer occurring in younger age groups. Lost potential years can be interpreted as lost productive years (both economic and non-economic) that a person dying prematurely of cancer would have contributed to society if he or she had survived. A person dying of cancer at age 35 years would have had 30 more years of potential life lost than a person dying of cancer at age 65.

Race/Ethnicity

The data used in the Race/Ethnicity section of the overview do not include those cases with race or ethnicity listed as “unknown”. The “Unknown” race category is the second most frequently reported after “white” and is the second most frequently reported Hispanic origin category after “Non-Hispanic”. There are ongoing registry efforts to improve the quality of race and ethnicity reporting from hospitals.

The incidence rates calculated by race use a different population denominator than the overall incidence rates presented in this report. See the *Data Sources* section for additional information.

Software

All incidence and mortality counts, including counts used for analysis of stage at diagnosis, crude and age-adjusted rates, and current (5-year) trends were generated using Surveillance Research Program, National Cancer Institute SEER*Stat software (www.seer.cancer.gov/seerstat) version 5.0.17. Data were formatted for SEER*Stat using Surveillance Research Program, National Cancer Institute SEER*Prep software version 2.1. Historical, six-year trends were calculated using Joinpoint Regression Software (www.srab.cancer.gov/joinpoint) Version 2.7, September 2003, National Cancer Institute. Smoothed rates for regional maps were calculated using Hansen Simonson and Surveillance Research Program, NCI Headbang software (http://srab.cancer.gov/headbang) Version 3.0.

Trends

All trends are calculated using age-adjusted rates and are reported as an annual percent change (APC). The APC is calculated by fitting a weighted, least squares regression line to the natural logarithm of the rates, using year as a regressor variable.

Current Five-Year Trends – These trends are calculated using two-year averages of the age-adjusted rates as endpoints. The purpose of these trends is to allow comparison of general Oregon trends with national trends based on direction (increase or decrease) and slope (rapid or slow change). This trend analysis is intended to describe broad, temporal changes of cancer rates in Oregon.

Historical Trends – Historical trends are calculated using age-adjusted rates for all years of data (1996-2001, six years, for this report). Because of limitations due to a small number of years of data, analysis was done assuming no joinpoints (constant APC percent change for all years) compared to results using one joinpoint to determine if any change in trends is statistically significant. The number of joinpoints considered statistically significant is determined using the permutation test. In subsequent years, the number of joinpoints analyzed will increase with increasing years of data available for analysis.

The purpose of these trends is to determine if there are any recent changes in Oregon trends. This trend analysis is intended to describe temporal changes in Oregon cancer rates with greater precision.