
Cervical Cancer Core Data Set

DATA COORDINATING COMMITTEE



**Indiana Cancer
Consortium**

October 2006

Data Coordinating Committee Members

Greg Steele, Co-chair

Indiana University School of Medicine, Department of Public Health

Mindy Hightower King, Co-chair

Indiana University, Center for Evaluation and Education Policy

Bob Chloupek

Private Physician

Leslie Dufner

American Cancer Society

Martha Graves

Indiana State Department of Health

Elizabeth Hamilton-Byrd

Indiana State Department of Health

John Lewis

Health Care Excel, Inc.

Michelle Lucke

IU Cancer Center

Miranda Spitznagle

Indiana Tobacco Prevention and Cessation

Linda Stemnock

Indiana State Department of Health

Kevin Turcotte

Ball State University, Department of Geography

Melissa Elrod

Comprehensive Cancer Control (Staff)

Amanda Hewitt

Comprehensive Cancer Control (Staff)

Tom Rich

Comprehensive Cancer Control (Staff)

William Thornton

Comprehensive Cancer Control (Staff)

Introduction

The Indiana Cancer Consortium (ICC) is proud to present this data set detailing cervical cancer incidence, mortality, histology groups, and screening rates for Indiana residents. The core datasets are the product of the ICC Data Committee for the use of corresponding ICC Coordinating Committees and Member Organizations. The committee is working toward achieving the goal in the *Indiana Cancer Control Plan 2005-2008* to “Increase the quantity, quality and availability of complete and timely cancer and related data.”

Comprehensive, timely, and accurate cancer-related data are essential for evaluating progress toward cancer prevention and control. Staging data and incidence, mortality, and survival rates help identify populations at greater risk for developing or dying of cancer, monitor epidemiological trends, and provide information for evaluating progress toward goals. Data make it possible for the ICC and its committees to review the current cancer burden in the state, set priorities, and assist with cancer control planning. In addition to gathering cancer data for short, intermediate, and long term evaluation of the goals and objectives of the cancer plan, specific efforts are focused on improving access to and increased usability of cancer and related data in Indiana.

Sources of data used to compile the datasets include the Indiana State Cancer Registry, the Indiana State Department of Health Epidemiology Resource Center, the Indiana Behavioral Risk Factor Surveillance System (BRFSS), and the National Cancer Institute [Surveillance Epidemiology and End Results](#) (SEER) System. A complete source of references is included following the data tables along with a listing of data limitations and an interpretation of the confidence intervals. Please note that while confidence intervals are listed only in certain tables, these are available for all tables upon request to admin@indianacancer.org.

Datasets for other cancers will be developed and published by the committee in the upcoming year.

Data Limitations

The findings from the BRFSS are subject to at least four limitations. First, responses are self reported and have not been confirmed by medical record review. Second, BRFSS is a telephone survey and does not cover persons without land-line telephones, persons in the military, or those residing in institutions. Telephone coverage in Indiana for 2002 is estimated as 94.67% and varies across counties, as well as for subgroups. For example, minorities and those in lower socioeconomic groups typically have lower telephone coverage; in addition, many younger adults rely exclusively on cell phones. Third, comparisons of the data presented by race/ethnicity or other important variables are unadjusted for age. The age distribution of a population (i.e., the number of people in particular age categories) can change over time, and can be different in different geographic areas and for different subgroups.

While the Indiana BRFSS data are not age-adjusted, the survey responses are weighted. The number of different telephone numbers that reach each household, the number of adults in each household, and the number of completed interviews in each cluster are considered in the weighting process. After adjusting the raw data to these three factors, the data are adjusted further using the Indiana age-race-sex group distribution so that the weighted sample data produce demographic distributions that correspond closely to the state population. Finally, while the Council of American Survey Research Organizations (CASRO) response rates for the BRFSS have decreased in recent years, the rates are considered satisfactory by CDC to yield valid data (Indiana's CASRO rate was 49.5% in 2004).

Mortality data are also subject to several limitations. First, death rates for Hispanic, AIAN (American Indian, Alaskan Native), and API (Asian/Pacific Islander) persons should be interpreted with caution because of inconsistencies in reporting Hispanic origin or race on the death certificate as compared with race or Hispanic origin on censuses, surveys, and birth certificates. Studies have shown underreporting on death certificates of AIAN, API, and Hispanic decedents; and under-counts of these groups in the censuses. A number of studies have been conducted on the reliability of race reported on the death certificate by comparing race on the death certificate with that reported on another data collection instrument, such as the census or a survey. Differences may arise because of differences in who provides race information on the compared records. Race information on the death certificate is reported by the funeral director as provided by an informant or in the absence of an informant, on the basis of observation. Second, cause of death statistics are affected by the accuracy of the certifier, who may not be familiar with the patient's medical history; by whether standard terminology is used by the certifier; by the legibility of the certifier's handwriting; and by the accuracy of the coder assigning cause of death codes.

Similarly, there are limitations to Cancer Incidence data. Although state registries across the country use standardized data items and codes for race and ethnicity (i.e., Hispanic origin), the initial collection of this information by health care facilities and practitioners and the procedures for assigning and verifying codes for race and ethnicity are not well standardized. Thus, some inconsistency is expected in this information. The abstraction and coding of cancer data is dependent on the medical record which may be incomplete, disorganized, or in some cases inaccurate. The completeness of cancer data is dependent upon the ascertainment of cases and timely reporting by individual facilities and practitioners. Overall the completeness and accuracy of the data of the Indiana State Cancer Registry is high, meeting standards for certification by the North American Association of Central Cancer Registries; however, completeness may be lower for some subgroups of the population. Comparisons among different subgroups of the population for Stage at Diagnosis

may be complicated by differing percentages of unstaged cases. The percent of cases unstaged varies not only by primary site, but also by race/ethnicity and age, and may also vary by primary payer and geography.

The accuracy of rates are dependent not only on the numerator (the number of cases diagnosed, the number of deaths, or the number of people answering “yes” to a question), but also on the denominator (the number of individuals in the population of interest). Numerator errors may be due to underreporting or to misclassification of items, such as age, race/ethnicity, cancer type, or cause of death. Denominator errors arise through under- and overenumeration in the decennial census, which is the basis of intercensal population estimates and population projections. To the extent that any over- or undercount is substantial and variable among subgroups or geographic areas, it may have important consequences on calculated death rate statistics. The effect of an *undercount* of population is that it decreases the denominator, leading to an *overestimate* of the rate. Conversely, an *overcount* of population would result in an *underestimate* of the rate.

Interpretation of Confidence Intervals

The confidence interval is a range of plausible values for a population parameter (measure) that has been estimated from a given set of data. In the case of cancer data, the parameters of interest are often rates. We have provided 95% confidence intervals for many of the rates in the core data set. The end points of the confidence interval are designated the lower confidence limit (LCL) and the upper confidence limit (UCL). The confidence interval reflects the amount of variability in the estimation of the cancer rate - the more variable the data, the wider the confidence interval. Similarly, the size of the population will affect the width of the confidence interval. For a particular parameter, a very small population will generally have a wider confidence interval than will a large population. The level of confidence chosen also affects the width of the interval, with a higher level of confidence that the true value of the parameter is included within the interval corresponding to a wider interval. Confidence intervals can be calculated at many different confidence levels. Some frequently chosen confidence levels are 90%, 95%, and 99%.

The confidence level is not the same as a significance level used to test for statistical significance. However, the confidence interval can be used as a rough guide to whether the true value of a parameter for one population (e.g., all men) is likely to be different from the true value of that parameter for another population (e.g., all women). If the two intervals do not overlap, then it is likely that the difference between the estimates of the parameters corresponds to an actual difference between the true values of the parameter in the two populations. If the LCL of one confidence interval overlaps the UCL of the other interval, there is unlikely to be a difference between the true values of the parameter for those populations. When the difference between the estimates of a parameter for two populations is large, a test of hypothesis could be used as a more sensitive method than the confidence interval of detecting a true difference between the two populations.

INCIDENCE RATES (95% Confidence Interval)

Table 1

	1998-2002				2002				2001			
	US		IN		US		IN		US		IN	
	#	Rate	CI	#	Rate	CI	#	Rate	CI	#	Rate	CI
Total Population	N/A	8.0	7.8-8.2	1,417	9.1	8.6-9.6	N/A	7.2	N/A	274	8.7	7.7-9.8
Total White	N/A	7.5	N/A	1,249	8.8	8.4-9.3	N/A	6.8	N/A	224	8.6	7.6-9.8
Total African-American	N/A	11.5	N/A	144	12.1	10.2-14.3	N/A	10.3	N/A	26	10.7	7.0-15.9

Sources: Indiana Cancer Incidence and Mortality Annual Report 2002 & 2003; Surveillance Epidemiology & End Results (SEER) 2004.

Notes: See attached reference page for complete reference.

MORTALITY RATES (95% Confidence Interval)

Table 2

	1998-2002				2002				2001			
	US		IN		US		IN		US		IN	
	#	Rate	CI	#	Rate	CI	#	Rate	CI	#	Rate	CI
Total Population	N/A	2.8	2.7-2.8	440	2.7	2.4-3.0	N/A	2.5	N/A	80	2.4	N/A
Total White	N/A	2.5	N/A	383	2.6	2.3-2.8	N/A	2.3	N/A	67	2.2	1.7-2.8
Total African American	N/A	5.3	N/A	53	4.8	3.6-6.3	N/A	5.0	N/A	12*	5.3*	2.7-9.4*

Sources: Indiana Cancer Incidence and Mortality Annual Report 2002 & 2003; Surveillance Epidemiology & End Results (SEER) 2004.

Notes: See attached reference page for complete reference; *<20 cases, unstable rate.

STAGE AT DIAGNOSIS

Table 3

	1995-2001				1997-2001				1998-2002						
	US		IN		US		IN		US		IN				
	T	L	R	D	Unk	T	L	R	D	Unk	T	L	R	D	Unk
All Races (T)	7,343	N/A	N/A	N/A	N/A	1,454	784	480	113	77	1,417	747	478	107	85
	100%	55.0%	32.0%	8.0%	6.0%	100%	53.9%	33.0%	7.8%	5.3%	100%	52.7%	33.7%	7.6%	6.0%
White	5,428	N/A	N/A	N/A	N/A	1,281	702	413	104	62	1,249	669	418	96	66
	100%	56.0%	30.0%	8.0%	5.0%	100%	54.8%	32.2%	8.1%	4.8%	100%	53.6%	33.5%	7.7%	5.3%
African-American	1,076	N/A	N/A	N/A	N/A	149	70	59	8	12	144	70	53	10	11
	100%	48.0%	35.0%	9.0%	8.0%	100%	47.0%	39.6%	5.4%	8.1%	100%	48.6%	36.8%	6.9%	7.6%

Source: Indiana Cancer Incidence and Mortality Annual Report 2002 & 2003.

Note: See attached reference page for complete reference; percentages may not add to 100% due to rounding.

T=Total; L=Local (does not include in situ); R=Regional; D=Distant; Unk=Unknown

MAJOR HISTOLOGY GROUPS (Indiana Data)

Table 4

Description	Count	% of Total
Squamous Cell Neoplasms	1,014	71.5%
Adenomas & Adenocarcinomas	235	16.7%
All Other	169	12.0%
Total	1,418	100.2%*

Source: Indiana State Cancer Registry Incidence Files (1998-2002).

Note: See attached reference page for complete reference; numbers do not sum to 100 due to rounding.

SCREENING DATA (Indiana Data)

Table 5 A Pap test is a test for cancer of the cervix. Have you ever had a Pap test?

Age Group	2004			2002			2000			
	% Yes	95% CI	% No	95% CI	Total N	% Yes	95% CI	% No	95% CI	Total N
18-39	93.1	91.2-95.1	6.9	4.9-8.8	1,249	92.5	90.3-94.6	7.5	5.4-9.7	1,158
40-64	98.1	97.4-98.9	1.9	1.1-2.6	1,755	98.9	98.3-99.4	1.1	0.6-1.7	1,552
65+	91.3	89.2-93.5	8.7	6.5-10.8	895	92.1	89.9-94.3	7.9	5.7-10.1	706
Race/Ethnicity										
White, Non-Hispanic	95.2	94.2-96.2	4.8	3.8-5.8	3,414	96.4	95.6-97.2	3.6	2.8-4.4	3,029
Black, Non-Hispanic	92.5	88.8-96.2	7.5	3.8-11.2	236	95.5	90.6-100.0	4.5	0.0-9.4	185
Other/Multiracial	92.6	87.2-98.0	7.4	2.0-12.8	104	81.3	71.8-90.8	18.7	9.2-28.2	101
Hispanic	92.5	87.2-97.8	7.5	2.2-12.8	145	77.2	66.3-88.0	22.8	12.0-33.7	101
Grand Totals	94.7	93.8-95.7	5.3	4.3-6.2	3,938	94.9	94.0-95.9	5.1	4.1-5.0	3,457

Source: Indiana Behavioral Risk Factor Surveillance System (BRFSS) 2000, 2002, 2004.

Notes: See attached reference page for complete reference; percents are based on weighted data; unless otherwise specified, totals are unweighted numbers.

***Asterisks indicates less than 50 total respondents; denominator is females who have had a Pap test.

SCREENING DATA (Indiana Data)

Table 6a How long has it been since you had your last Pap test?

Age Group	2004					2002					2000					
	% Within the past year	95% CI	% 1 year to <2 years ago	95% CI	% 2 years to <3 years ago	95% CI	% 3 years to <5 years ago	95% CI	% 5 or more years ago	95% CI	Total N	% Yes	95% CI	% No	95% CI	Total N
18-39	72.3	69.5-75.2	16.1	13.8-18.5	4.6	3.4-5.8	3.1	2.0-4.1	3.8	2.6-5.1	1,184	92.9	90.1-95.8	7.1	4.2-10.0	638
40-64	56.8	54.2-59.3	17.2	15.3-19.2	7.4	6.1-8.8	6.7	5.3-8.0	11.9	10.2-13.5	1,712	97.5	95.7-99.4	2.5	0.6-4.3	670
65+	33.0	29.4-36.7	16.8	13.8-19.7	12.2	9.7-14.8	9.6	7.3-12.0	28.3	24.8-31.9	784	89.6	86.3-92.9	10.4	7.1-13.7	381
Race/Ethnicity																
White, Non-Hispanic	57.6	55.7-59.5	16.7	15.2-18.1	7.2	6.3-8.2	6.1	5.1-7.0	12.4	11.2-13.6	3,235	94.2	92.6-95.8	5.8	4.2-7.4	1,532
Black, Non-Hispanic	65.7	58.3-73.0	18.0	12.2-23.7	6.5	2.8-10.3	4.8	1.8-7.9	5.0	2.2-7.7	218	96.4	93.2-99.7	3.6	0.3-6.8	114
Other/Multiracial	62.4	51.5-73.4	16.0	8.1-23.9	5.4	1.1-9.6	3.1	0.0-6.4	13.1	4.8-21.3	95	*	*	*	*	21
Hispanic	65.1	56.5-73.8	18.4	11.4-25.4	7.2	3.0-11.4	3.6	0.0-7.3	5.7	2.2-9.3	134	*	*	*	*	36
Total Female	58.5	56.7-60.3	16.8	15.4-18.1	7.2	6.3-8.1	5.8	5.0-6.6	11.7	10.6-12.8	3,712	94.1	92.6-95.6	5.9	4.4-7.4	1,708

Source: Indiana Behavioral Risk Factor Surveillance System (BRFSS) 2000, 2002, 2004.

Notes: See attached reference page for complete reference; percents are based on weighted data; unless otherwise specified, totals are unweighted numbers.

***Asterisks indicates less than 50 total respondents; denominator is females who have had a Pap test.

SCREENING DATA (Indiana Data)
Table 6b

How long has it been since you had your last Pap test?

2002										
Age Group	% Within the past year	95% CI	% 1 year to <2 years ago	95% CI	% 2 years to <3 years ago	95% CI	% 3 years to <5 years ago	95% CI	% 5 or more years ago	Total N
18-39	76.9	74.1-79.7	15.0	12.6-17.5	3.4	2.3-4.5	1.9	1.1-2.7	2.8	1,101
40-64	61.8	59.0-64.6	15.0	13.0-17.0	6.5	5.0-8.0	4.3	3.1-5.4	12.5	1,518
65+	40.0	35.6-44.4	16.8	13.5-20.2	7.9	5.5-10.4	7.6	5.3-9.9	27.6	624
Race/Ethnicity										
White, Non-Hispanic	63.1	61.1-65.0	15.5	14.0-16.9	5.4	4.5-6.5	4.0	3.2-4.7	12.0	2,894
Black, Non-Hispanic	70.6	62.7-78.6	14.2	7.4-21.1	6.6	3.3-9.9	1.9	0.0-4.1	6.6	178
Other/Multiracial	64.0	52.4-75.6	11.1	3.3-18.8	7.7	0.8-14.7	5.5	0.6-10.4	11.6	86
Hispanic	70.5	59.9-81.2	14.0	5.7-22.4	4.4	0.2-8.5	4.1	0.1-8.2	6.9	85
Total Female	63.8	61.9-65.6	15.3	13.9-16.7	5.6	4.7-6.4	3.9	3.2-4.6	11.4	3,277

Source: Indiana Behavioral Risk Factor Surveillance System (BRFSS) 2000, 2002, 2004.

Notes: See attached reference page for complete reference; percents are based on weighted data; unless otherwise specified, totals are unweighted numbers.

***Asterisks indicates less than 50 total respondents; denominator is females who have had a Pap test.

SCREENING DATA (Indiana Data)
Table 6c

How long has it been since you had your last Pap test?

2000										
Age Group	% Within the past year	95% CI	% 1 year to <2 years ago	95% CI	% 2 years to <3 years ago	95% CI	% 3 years to <5 years ago	95% CI	% 5 or more years ago	Total N
18-39	75.7	71.8-79.7	12.8	9.8-15.7	6.2	3.8-8.7	1.7	0.5-2.9	3.5	606
40-64	66.5	62.5-70.5	14.1	11.2-17.1	6.8	4.4-9.3	3.5	2.0-4.9	9.1	655
65+	41.5	35.7-47.3	16.1	11.8-20.4	7.4	4.2-10.5	7.1	4.1-10.2	27.8	328
Race/Ethnicity										
White, Non-Hispanic	64.6	61.9-67.4	13.9	11.9-15.9	6.7	5.1-8.2	3.8	2.7-4.8	11.1	1,443
Black, Non-Hispanic	66.0	55.8-76.3	16.9	8.6-25.2	7.5	1.2-13.8	3.2	0.0-7.1	6.3	107
Other/Multiracial	*	*	*	*	*	*	*	*	*	16
Hispanic	*	*	*	*	*	*	*	*	*	35
Total Population	65.5	62.9-68.1	13.9	12.1-15.8	6.6	5.1-8.1	3.6	2.6-4.5	10.4	1,605

Source: Indiana Behavioral Risk Factor Surveillance System (BRFSS) 2000, 2002, 2004.

Notes: See attached reference page for complete reference; percents are based on weighted data; unless otherwise specified, totals are unweighted numbers.

***Asterisks indicates less than 50 total respondents; denominator is females who have had a Pap test.

Cervical Cancer Core Data Set References

Tables 1-3

Indiana State Cancer Registry and the Epidemiology Resource Center (July 2005). **Cancer Incidence and Mortality 2002 Annual Report**. Retrieved January 9, 2006 from <http://www.in.gov/isdh/dataandstats/cancer.htm>.

Indiana State Cancer Registry and the Epidemiology Resource Center (March 2005). **Cancer Incidence and Mortality 2001 Annual Report**. Retrieved January 9, 2006 from <http://www.in.gov/isdh/dataandstats/cancer.htm>.

Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, Mariotto A, Feuer EJ, Edwards BK (eds). *SEER Cancer Statistics Review, 1975-2002*, National Cancer Institute, Bethesda, MD. http://seer.cancer.gov/csr/1975_2002, based on November 2004 SEER data submission, posted to the SEER web site 2005.

Note: The United States incidence and mortality rates for Whites and African-Americans are from SEER source above.

Table 4

Indiana State Department of Health, Indiana State Cancer Registry Incidence Files (1998-2002), and the Epidemiology Resource Center (March 2006).

Notes: Rocky Mountain Cancer Data Systems (FORDS system, version 4.03) was used to produce the histology data. The subset (cvx9802.sel) was created on 2/22/2006. Lymphomas were not excluded from the criteria, and there is one lymphoma of the cervix in the data provided in Table 4. The following variables were used to produce the histology data:

ZIP Code at diagnosis = 460000000 to 479999999 *or* State at diagnosis = Indiana

Sex = Female

Site = ICD-O-3 C53.0-C53.9 (Cervix Uteri)

Date of diagnosis = 00/00/1998 to 99/99/2002

Tables 5 & 6

Indiana State Department of Health. **Indiana Behavioral Risk Factor Surveillance System** (2000). Retrieved January 9, 2006 from http://www.in.gov/isdh/dataandstats/brfss/brfss_index.htm.

Indiana State Department of Health. **Indiana Behavioral Risk Factor Surveillance System** (2002). Retrieved January 9, 2006 from http://www.in.gov/isdh/dataandstats/brfss/brfss_index.htm.

Indiana State Department of Health. **Indiana Behavioral Risk Factor Surveillance System** (2004). Retrieved January 9, 2006 from http://www.in.gov/isdh/dataandstats/brfss/brfss_index.htm.

Notes: For the age groups in Tables 5 and 6, 13 five-year age groups were combined into three groups: 18-39, 40-64, and 65+. The two cervical cancer screening questions were run using the three age groupings and using the state weight for calculating the data. Responses for the unknown and refused responses to the age question were suppressed because the number was less than 50.

* Consult the table references for specific methods.