

Field Epidemiology Training Program

Cancer Curriculum: Principles of Cancer Registries

**In-class Exercise:
Assessing Data Quality**

PARTICIPANT GUIDE



Elsewhere Population-Based Cancer Registry, 2003–2006

You have been given data from a population-based cancer registry in the state of 'Elsewhere' in your country. The data covers incident cancers registered in 2003–2006. Your goal is to assess the quality of these data and make a determination about whether the data should be published in the next volume of *Cancer Incidence in Five Continents (CI5)*. Your assessment will focus primarily on validity and completeness of the data.

Along with the data, the registry provided the following background information:

- A list of all data sources used to generate the data—these included public and private hospitals, pathology labs, death certificates from hospitals and the local vital registration office and numerous primary care clinics. **You determine that the diversity and number of sources appears to be appropriate.**
- A description of data collection practices, adherence to internationally accepted rules for tumor classification and calculation of incident cancers. Based on this information, **you determine that the data are likely to be comparable** across years of data collected and to other registries.
- A description of consistency checks routinely performed by registry staff using CanReg5

The cancer registry also submitted population census data, including population counts by sex and 5 year age group. These were used to calculate crude and adjusted estimates of cancer incidence by site using the world standard population.

Review Editorial Table 5, the population pyramid for Elsewhere:

1. Does the population appear to be expanding, stable or contracting?
2. Assuming cancer incidence rates remain the same in this population, do you expect the total number of new cancer cases per year to increase, decrease or remain about the same?

Next review **Editorial Table 1**: “Number of cases in major diagnosis groups in single calendar years of observation” and “ASR in major diagnosis groups in single calendar years of observation” for males, females and both sexes in 2003 – 2006 (Note that the numbers in parenthesis represent % of total count or ASR in given year).

Review the tables and bar charts for “All sites but skin – Both sexes” for each year.

3. Are the total numbers and % of total for “All sites but skin” stable from one year to the next for males, females and both sexes? Describe any major differences you observe.

4. Are the ASR and % of total for “All sites but skin” stable from one year to the next for males, females and both sexes? Describe any differences you observe.

5. What might account for the differences you observed in counts and ASR in 2003 – 2006?

6. What does this suggest about the completeness of case capture?

Editorial table 2 provides graphs of age-specific rates for the major cancer sites. Incidence rates are plotted on the y-axis by age on the x-axis. Males are shown in blue and females are shown in red. The data are plotted for males and females on a logarithmic scale on the second page.

Compare the shape of the age-specific rates for each cancer site to the reference tables provided for each cancer site in the appendix. Focus on the shape of the age-specific rates, rather than the incidence rates themselves.

7. Do you observe any differences in age-specific incidence rates between males and females by anatomic location? What real or artefactual explanations could account for any differences you observe?

Finally, look at **Editorial Table 4**: “International Comparison and M/I Ratios”, which contains the number of cases, age-standardized rate (ASR), observed/expected incidence rate (O/E), percent microscopically verified (MV%), percent death certificate only (%DCO) and mortality:incidence ratio (M/I%) for each of the major cancer sites listed.

In this section you will compare the observed values to the regional indicators provided in the appendix. Typically, statistical tests would be applied to look for significantly higher or lower values in the registry’s data as compared to the regional standard. For this exercise, we would like you to focus on any major differences you observe solely upon visual inspection of the data.

8. Look at the number of cases registered per cancer site. Based on what you see, do you have any concerns about the stability of the ASR for any of the cancer sites?

9. Focusing only on cancer sites for which there are at least 20 cases, review the O/E ASR for each cancer site for males and females. For the purposes of this evaluation, an O/E substantially different than 1.0 (i.e. <0.7 or >1.3) warrants further consideration.
 - a. Choose one cancer site in males or females in which the O/E is substantially >1.0. What might explain this discrepancy?

- b. Choose one cancer site in males or females in which the O/E is substantially <1.0 . What might explain this discrepancy?

10. Review the %MV for each cancer site for both sexes. How well do these align with the indicators for the region?

- a. What does this tell you about the validity of the data?

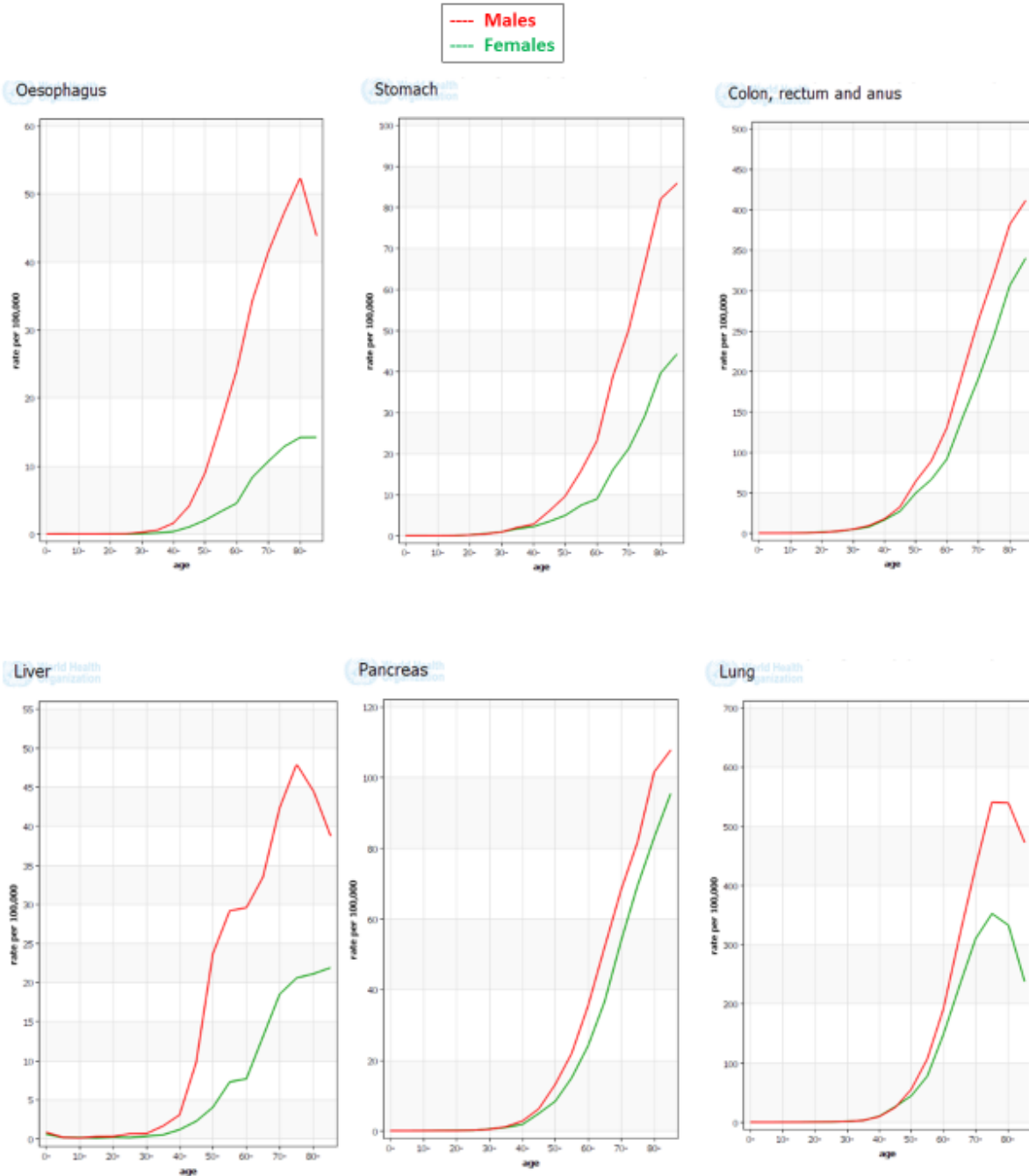
- b. What does this suggest about the completeness of the data?

11. Are the cancers diagnosed by DCO those you might expect to be diagnosed by DCO in some cases?

- a. Based on the %DCO, what would you expect to see if you reviewed the data for % of missing data?
-
12. Compare the M/I% to the regional standards provided. Do any M/I% values stand out as potentially concerning for completeness of the data?
 - a. What are possible explanations for M/I% values that are >100%?
-
13. Based on your assessment of validity and completeness, would you recommend publication of these data in the next volume of CI5? Please justify your answer.

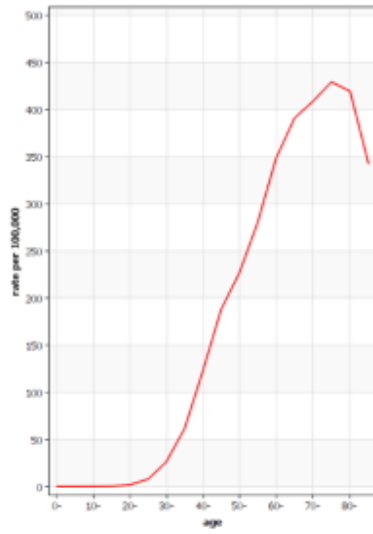
Appendix

Reference age-specific incidence rates graphs for major cancer sites, 2003–2007

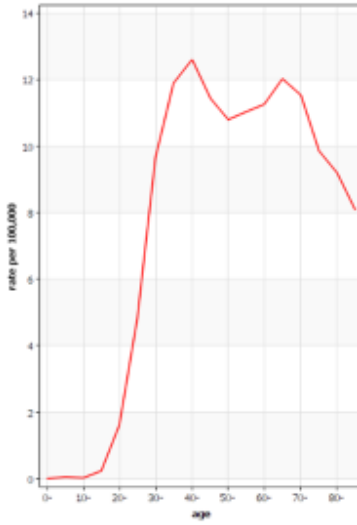


FETP Cancer Curriculum: Principles of Cancer Registries
In-class Exercise: Assessing Data Quality

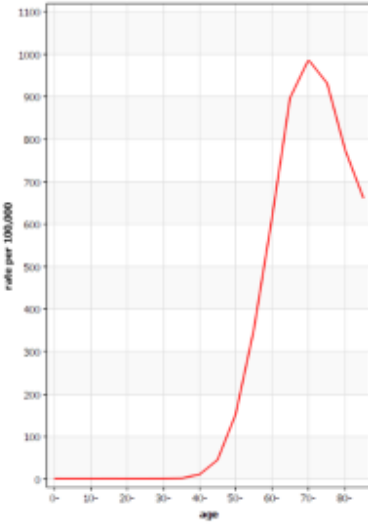
Breast Health Statistics



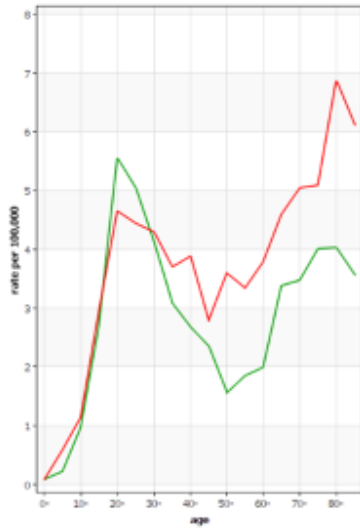
Cervix uteri



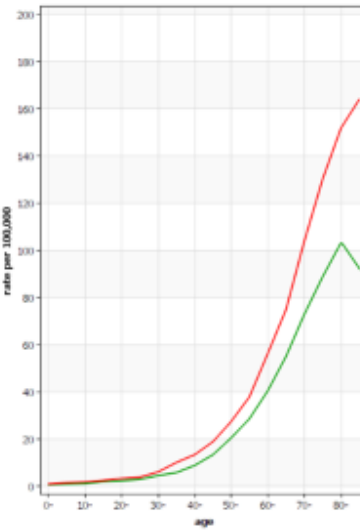
Prostate Health Statistics



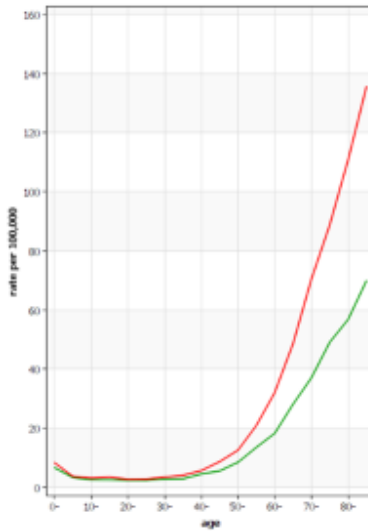
Hodgkin lymphoma



Non-Hodgkin lymphoma



Leukaemia



Mean values of data quality indicators for cancer registries in the Elsewhere region

ICD-10 code	Cancer site	Male			Female		
		MV%	M:I%	ASR	MV%	M:I%	ASR
C00–14	Oral cavity and pharynx	94.2	44.0	12.2	92.8	42.0	3.4
C15	Oesophagus	81.3	99.8	5.3	73.5	89.3	1.1
C16	Stomach	77.4	75.9	18.3	73.5	76.8	8.9
C18–21	Large bowel	85.3	49.9	17.5	82.7	49.5	17.3
C22	Liver	39.7	74.5	4.6	31.5	84.8	2.8
C25	Pancreas	39.9	75.3	4.3	34.0	79.1	3.3
C32	Larynx	91.9	57.2	6.4	87.4	58.3	0.9
C33–34	Trachea, bronchus, and lung	60.0	95.7	21.1	57.3	101.2	8.3
C43	Melanoma of skin	98.6	35.7	1.6	100.0	30.2	1.2
C50	Breast	92.8	27.9	0.5	95.0	31.6	42.9
C53	Cervix uteri	0.0	0.0	0.0	90.5	35.5	13.3
C54–55	Corpus uteri, uterus unspecified	0.0	0.0	0.0	89.1	33.4	8.3
C56	Ovary	0.0	0.0	0.0	80.9	37.8	5.6
C61	Prostate	89.0	32.9	79.9	0.0	0.0	0.0
C62	Testis	90.9	11.7	1.6	0.0	0.0	0.0
C64–66	Kidney, renal pelvis, and ureter	78.3	32.3	3.6	79.1	30.4	2.3
C67	Bladder	88.7	30.0	8.1	83.9	37.1	2.3
C70–72	Brain, central nervous system	72.8	53.5	4.3	68.1	55.3	2.9
C73	Thyroid	95.9	9.1	1.3	95.2	5.9	6.7
C81–88, C90	Lymphomas	95.0	53.6	12.2	94.9	53.2	8.9
C91–95	Leukaemia	85.5	77.7	7.1	81.4	71.3	4.9
C76–80	Unspecified	47.7	54.4	6.5	48.5	52.2	5.2
C00–96 (excluding C44)	All sites (excluding non-melanoma skin)	81.1	51.9	226.7	82.3	46.8	159.3