THE PROCESS OF CANCER REGISTRY DATA QUALITY EVALUATION

Nadya Dimitrova
OUTLINE

- Introduction
- Comparability
- Completeness
- Validity
- Timeliness
Cancer registry (CR) is:

- an information system
- collection
- storage
- management
- analysis of data on persons with cancer (general or specialized CR)
- covering a whole country or a region (population-based) or hospitals/pathological laboratories (clinical/pathological CR).

History:

- 1926: First cancer registry in Hamburg
- 1942: Danish Cancer Registry
- 1950-1960: Start of Europe-wide implementation of cancer registration
- Today: about 150 - 200 CR in Europe
about 160 cancer registries in Europe

eco.iarc.fr
ENCR membership survey 2014

1. They operate in Europe.
2. They completed the ENCR questionnaire 'Overview of the registration practices' launched on May 2010 and on May 2014.

159 members (37 countries and 66% of population)
(incl. specialized CRs)

21 countries (1 registry) ........Spain (16), France (20) and Italy (37)

In the EU: 20 national and 82 regional CRs, covering 72% of the population
2011 EUROCOURSE Survey on the Methods Used to Assess Completeness*

- Contacted: 179 European CRs
- Replied: 116 (65%)
- Of the respondent: 88% check completeness in some way
- At least one quantitative method: 53%
- Reasons for not estimating completeness:
  - Lack of:
    - Time
    - Software
    - Trained staff
  - Not necessary

* Zanetti R. et al. Completeness and timeliness: Cancer registries could/should improve their performance. EJC (2015), 51:1091-1098
Use of main methods for estimating completeness.*
European cancer registry by region (countries grouped according to the definition of the United Nations Population Division).

Zanetti R. Completeness and timeliness: Cancer registries could/should improve their performance. EJC (2015) 51:1091-1098
DATA QUALITY OF THE CR DATA – WHY IS IMPORTANT TO BE MONITORED

- To provide assurance that:
  - Incidence rates are accurate;
  - Survival estimates – unbiased;
  - Other statistics – reliable for use in research or decision making regarding cancer control
  - The CR data – considered ready for publication
DIMENSIONS OF DATA QUALITY AT CR

- **Comparability** - the statistics have to be comparable between different *populations* and over *time*.
- **Completeness** - the extent to which all diagnosed *neoplasms* in the population are included in the registry database.
- **Validity (accuracy)** - the proportion of cases in the CR with a given characteristic which *truly have* this attribute.
- **Timeliness** – rapidity at which a CR can collect, process and report sufficiently reliable and complete cancer data.
COMPARABILITY

- Statistics: comparable between different **populations** and over **time**.

- International **classifications** and **standards** for cancer registration:
  - ICD-O– topography, morphology, behavior, grade.
  - Recommendations for defining incidence date, basis of diagnosis, multiple primaries, etc.
  - TNM- and other classifications for staging

- Identified: **review** of the registration routines in place, specification of the standards and definitions
**EVALUATION OF COMPARABILITY - EXAMPLE**

- **Topics:**
  - The system used for *classification and coding* of neoplasms

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Morphology</strong></td>
<td>In-house 3-digit system 1953-1970</td>
<td>MOTNAC 4-digit system (with local modifications) 1970-1992</td>
<td>Non-solid tumours only: Kiel's classification with local extensions 1986-2001</td>
<td>Non-solid tumours only: ICD-O-3 with local extensions 2002-Present</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CONVERSION  The 3-digit system recoded to MOTNAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Topography</strong></td>
<td>In-house system similar to ICD-6 1953-1970</td>
<td>ICD-7 (with local modifications) 1970-1992</td>
<td></td>
<td>ICD-O-2 (with local modifications) 1993-present</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CONVERSION  All codes recoded to ICD-7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 1** – The standards of classification and coding of neoplasms followed, Norway 1953–2007.
EVALUATION OF COMPARABILITY - EXAMPLE

Topics:
- The definition of incidence – what is a case

Incident cases in Norway comprise all malignant and in situ neoplasms, and the incidence reported in Cancer in Norway (CiN) includes all cases with the 5th-digit behaviour code 3 according to ICD-O-3, for haematological malignancies, and ICD-O-2, for other tumours. Cases with 5th-digit behaviour code 1 are also included for tumours of the central nervous system.
The in-house rules for the registration of incidence date (for the reporting of new cases or calculating survival) depart from the European Network of Cancer Registries (ENCR) recommendations, as the Registry always registers the earliest incidence dates reported on the sources of notification, whereas the ENCR rules are based upon a hierarchy of possible sources. For the period 2001–2005, 19.7% of the cases had a different date on applying the ENCR rules. For these cases, the median difference between the ENCR-defined date and the in-house incidence date was 10 days.

For reporting and comparability with other registries it is, however, possible, when needed, to select and give priority to the date when the specimen was taken.
EVALUATION OF COMPARABILITY - EXAMPLE

Topics:
- The distinction between a primary cancer (new case) and an extension, recurrence or metastasis of an existing one

The recording of multiple primary tumours in the main follows the recommendations given by ENCR. The recognition of two or more primary cancers does not depend on time, and the groups of topography codes considered as single sites (from ICD-O-2 and ICD-O-3) are followed, with systemic and multicentric cancers counted only once. The CRN has, however, used a more detailed grouping of specific histologies constitute a new tumour or a recurrence. For the purposes of incidence reporting, the yearly publication from the CRN includes the first primary tumour within the same three character categories of the topography code in each patient.
COMPARABILITY - HINT

- Have all of these details written down in a Manual
- Don't rely only on the historical memory of the staff
COMPLETENESS

- The extent to which all diagnosed neoplasms in the population are included in the registry database.
- Maximum completeness - incidence rates and survival proportions close to their true values
EVALUATION OF COMPLETENESS

- **Qualitative (semi-quantitative) methods** – give an indication of the degree of completeness relative to other registries, or over time:

  - Historic data/comparative methods
    - **stability** of incidence rates over time;
    - **comparison** of incidence rates with other (similar) populations;
    - **shape** of age-specific curves;
    - incidence rates of **childhood cancers**
  - Mortality to Incidence ratio (M:I)
  - Number of sources/notifications per case
  - Microscopic verification of diagnosis
EVALUATION OF COMPLETENESS

Quantitative methods – provide a numerical evaluation of the extent to which all eligible cases have been registered:

- Independent case ascertainment;
- Capture-recapture methods;
- Death-certificate methods:
  - DCN/M:I
  - Flow method
EVALUATION OF COMPLETENESS - EXAMPLE

- Incidence rates of childhood cancers

**Table 1 – Age-specific incidence rates per 100,000 for childhood cancer by gender, Norway, 2001–2005.**

<table>
<thead>
<tr>
<th>Age</th>
<th>Boys</th>
<th>Reference</th>
<th>Girls</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–4</td>
<td>23.5</td>
<td>(12.3–24.7)</td>
<td>19.5</td>
<td>(9.7–21.4)</td>
</tr>
<tr>
<td>5–9</td>
<td>12.3</td>
<td>(8.5–15.6)</td>
<td>14.0</td>
<td>(6.9–12.0)</td>
</tr>
<tr>
<td>10–14</td>
<td>15.3</td>
<td>(8.5–15.0)</td>
<td>11.1</td>
<td>(6.8–13.6)</td>
</tr>
</tbody>
</table>

**Table 5.3. The lowest and highest deciles of incidence rates (per million) of childhood cancer in Volume IX**

<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>Boys Lowest</th>
<th>Boys Highest</th>
<th>Girls Lowest</th>
<th>Girls Highest</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–4</td>
<td>&lt; 13.7</td>
<td>&gt; 25.6</td>
<td>&lt; 11.3</td>
<td>&gt; 23.2</td>
</tr>
<tr>
<td>5–9</td>
<td>&lt; 8.9</td>
<td>&gt; 16.5</td>
<td>&lt; 7.0</td>
<td>&gt; 12.7</td>
</tr>
<tr>
<td>10–14</td>
<td>&lt; 9.2</td>
<td>&gt; 16.3</td>
<td>&lt; 7.9</td>
<td>&gt; 14.9</td>
</tr>
</tbody>
</table>
Data quality at the Bulgarian National Cancer Registry: An overview of comparability, completeness, validity and timeliness

Nadya Dimitrova, Donald Maxwell Parkin


<table>
<thead>
<tr>
<th>Cancer site</th>
<th>ICD10</th>
<th>Number of cases</th>
<th>Completeness estimates (%)</th>
<th>Capture-recapture</th>
<th>DCN/M:I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lip, oral cavity and pharynx</td>
<td>C00–C14</td>
<td>4410</td>
<td>97.2</td>
<td>94.5</td>
<td></td>
</tr>
<tr>
<td>Digestive organs</td>
<td>C15–25</td>
<td>42,965</td>
<td>95.6</td>
<td>94.4</td>
<td></td>
</tr>
<tr>
<td>Respiratory system</td>
<td>C30–34</td>
<td>22,788</td>
<td>98.3</td>
<td>97.5</td>
<td></td>
</tr>
<tr>
<td>Bones</td>
<td>C40–41</td>
<td>405</td>
<td>86.7</td>
<td>104.0</td>
<td></td>
</tr>
<tr>
<td>Soft tissues</td>
<td>C47, 49</td>
<td>907</td>
<td>98.5</td>
<td>90.5</td>
<td></td>
</tr>
<tr>
<td>Skin melanoma</td>
<td>C43</td>
<td>2111</td>
<td>97.1</td>
<td>95.4</td>
<td></td>
</tr>
<tr>
<td>Female breast</td>
<td>C50</td>
<td>18,547</td>
<td>94.7</td>
<td>95.3</td>
<td></td>
</tr>
<tr>
<td>Female genital organs</td>
<td>C51–58</td>
<td>17,186</td>
<td>96.6</td>
<td>95.1</td>
<td></td>
</tr>
<tr>
<td>Male genital organs</td>
<td>C60–63</td>
<td>9478</td>
<td>94.4</td>
<td>92.9</td>
<td></td>
</tr>
<tr>
<td>Urinary tract</td>
<td>C64–68</td>
<td>11,191</td>
<td>94.9</td>
<td>92.0</td>
<td></td>
</tr>
<tr>
<td>Central nervous system</td>
<td>C70–72</td>
<td>3481</td>
<td>94.1</td>
<td>94.8</td>
<td></td>
</tr>
<tr>
<td>Hematological malignancies</td>
<td>C81–96</td>
<td>7582</td>
<td>95.9</td>
<td>92.9</td>
<td></td>
</tr>
<tr>
<td>All sites</td>
<td>C00–C96</td>
<td>170,438</td>
<td>94.7</td>
<td>92.6</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>C00–C96</td>
<td>88,768</td>
<td>96.0</td>
<td>94.0</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>C00–C96</td>
<td>81,670</td>
<td>93.6</td>
<td>91.0</td>
<td></td>
</tr>
</tbody>
</table>
COMPLETENESS - HINT

- You don’t have to use all the methods
- Start with the simplest ones and choose those more appropriate for the CR data and procedures
Validity

The proportion of cases in a dataset with a given characteristic which truly have the attribute. It is estimated by:

- **Reabstracting and recoding** – going back to one or more sources, to check on accuracy of recording, or recoding exercises and correcting any obvious deficiencies;
- **Internal consistency** – check program (CanReg, IARCTools, ENCR-JRC software)
- **Morphological verification (MV%)** – percentage of cases with a morphologically verified diagnosis;
- **Death certificate only (DCO%)** – percentage of cases for which the only information came from a death certificate;
- **Missing information** – proportion (or percentage) of cases with missing data:
  - Primary site uncertain (PSU%) – C80, C26, C39, C48, C76
  - Age, sex
  - Stage

<table>
<thead>
<tr>
<th>Males</th>
<th>MV%</th>
<th>DCO%</th>
<th>PSU%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulgaria</td>
<td>73.3</td>
<td>9.8</td>
<td>5.0</td>
</tr>
<tr>
<td>Croatia</td>
<td>77.0</td>
<td>5.0</td>
<td>3.9</td>
</tr>
<tr>
<td>Estonia</td>
<td>89.2</td>
<td>-</td>
<td>2.5</td>
</tr>
<tr>
<td>Latvia</td>
<td>71.5</td>
<td>11.2</td>
<td>2.4</td>
</tr>
<tr>
<td>Lithuania</td>
<td>83.8</td>
<td>4.9</td>
<td>2.5</td>
</tr>
<tr>
<td>Serbia</td>
<td>78.3</td>
<td>6.5</td>
<td>2.9</td>
</tr>
<tr>
<td>Slovakia</td>
<td>85.6</td>
<td>9.0</td>
<td>2.3</td>
</tr>
<tr>
<td>Slovenia</td>
<td>91.9</td>
<td>0.9</td>
<td>2.4</td>
</tr>
<tr>
<td>average</td>
<td>81.3</td>
<td>6.8</td>
<td>3.0</td>
</tr>
</tbody>
</table>
Recent trends in cancer incidence: impact of risk factors, diagnostic activities and data quality of registration

Silvia Dehler¹, Simeon Tonev², Dimitri Korol¹, Sabine Rohrmann³, and Nadya Dimitrova²

¹Cancer Registry Zurich and Zug, Switzerland; ²National Cancer Registry, Bulgaria; ³Institute of Social and Preventive Medicine, University of Zurich, Zurich, Switzerland

Figure 2 - Annual percentage of death-certificate-only (DCO) and morphologically verified (MV) cases in the years 2000-2009 in Sofia and the Canton of Zurich.
TIMELINESS

- Rapidity at which a CR can collect, process and report sufficiently reliable and complete cancer data.

<table>
<thead>
<tr>
<th>ICD-10</th>
<th>Site</th>
<th>Cases 2005</th>
<th>Difference</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2006</td>
<td>2007</td>
<td></td>
</tr>
<tr>
<td>C00-95</td>
<td>All sites</td>
<td>24,199</td>
<td>24,730</td>
<td>531</td>
</tr>
<tr>
<td>C00</td>
<td>Lip</td>
<td>77</td>
<td>85</td>
<td>8</td>
</tr>
<tr>
<td>C01-02</td>
<td>Tongue</td>
<td>77</td>
<td>76</td>
<td>-1</td>
</tr>
<tr>
<td>C03-06</td>
<td>Mouth, other</td>
<td>81</td>
<td>81</td>
<td>0</td>
</tr>
<tr>
<td>C07-08</td>
<td>Salivary glands</td>
<td>48</td>
<td>51</td>
<td>3</td>
</tr>
<tr>
<td>C09-14</td>
<td>Pharynx</td>
<td>132</td>
<td>127</td>
<td>-5</td>
</tr>
<tr>
<td>C15</td>
<td>Oesophagus</td>
<td>198</td>
<td>197</td>
<td>-1</td>
</tr>
</tbody>
</table>
**TIMELINESS**

- **EUROCOURSE survey**

<table>
<thead>
<tr>
<th>Latency for completing 1 year of case ascertainment and releasing data of European cancer registries (in months).</th>
<th>n</th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete 1 year of incidence</td>
<td>113</td>
<td>21</td>
<td>18</td>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>Publish printed report</td>
<td>92</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>Publish data on internet</td>
<td>89</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>30</td>
</tr>
</tbody>
</table>
TIMELINESS

- CRs tend to delay publication of results waiting to achieve better completeness.
- A rational trade-off between completeness and timeliness:
  - Flow method – measuring completeness during the registration process
  - Short-term predictions to provide the estimates for the current year.
CONCLUSION

- All CRs **should be able** to provide some objective indication of the quality of the data that they have collected, describing:
  - Comparability
  - Completeness
  - Validity
  - Timeliness

- European CRs **could improve** their performance in DQ evaluation with the **support** of international registry groups, providing **recommendations and standardization of methods**.
REFERENCES

I am an imperfect person
Loved by a perfect God