The process of cancer registry data quality evaluation

Nadya Dimitrova

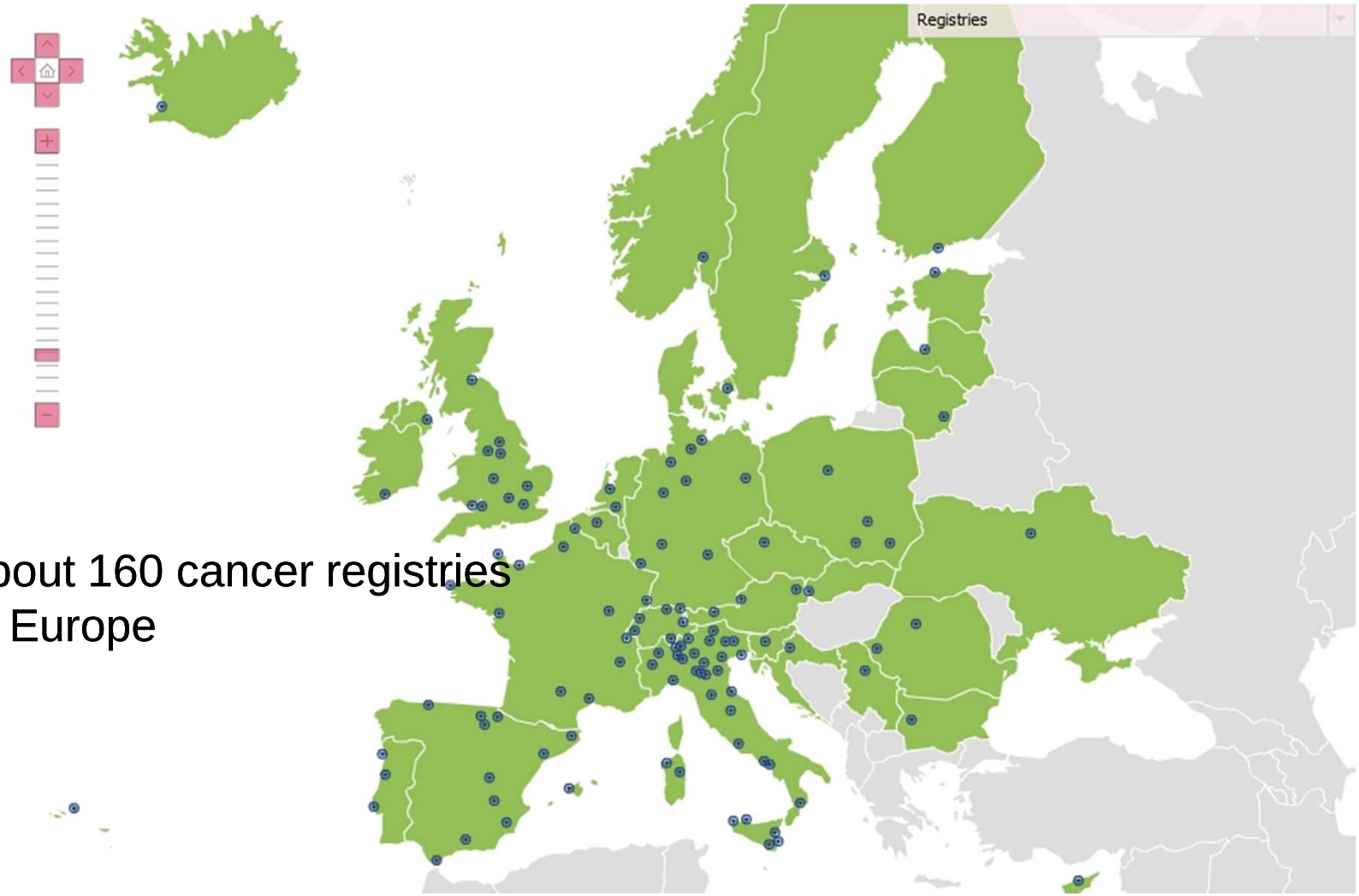
Outline

- Introduction
- Comparability
- Completeness
- Validity
- Timeliness

CANCER REGISTRATION IN EUROPE

- 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

European Network of Cancer Registries



about 160 cancer registries
in Europe

eco.iarc.fr

ENCR membership survey 2014

ENCR membership

Population-based cancer registries

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)

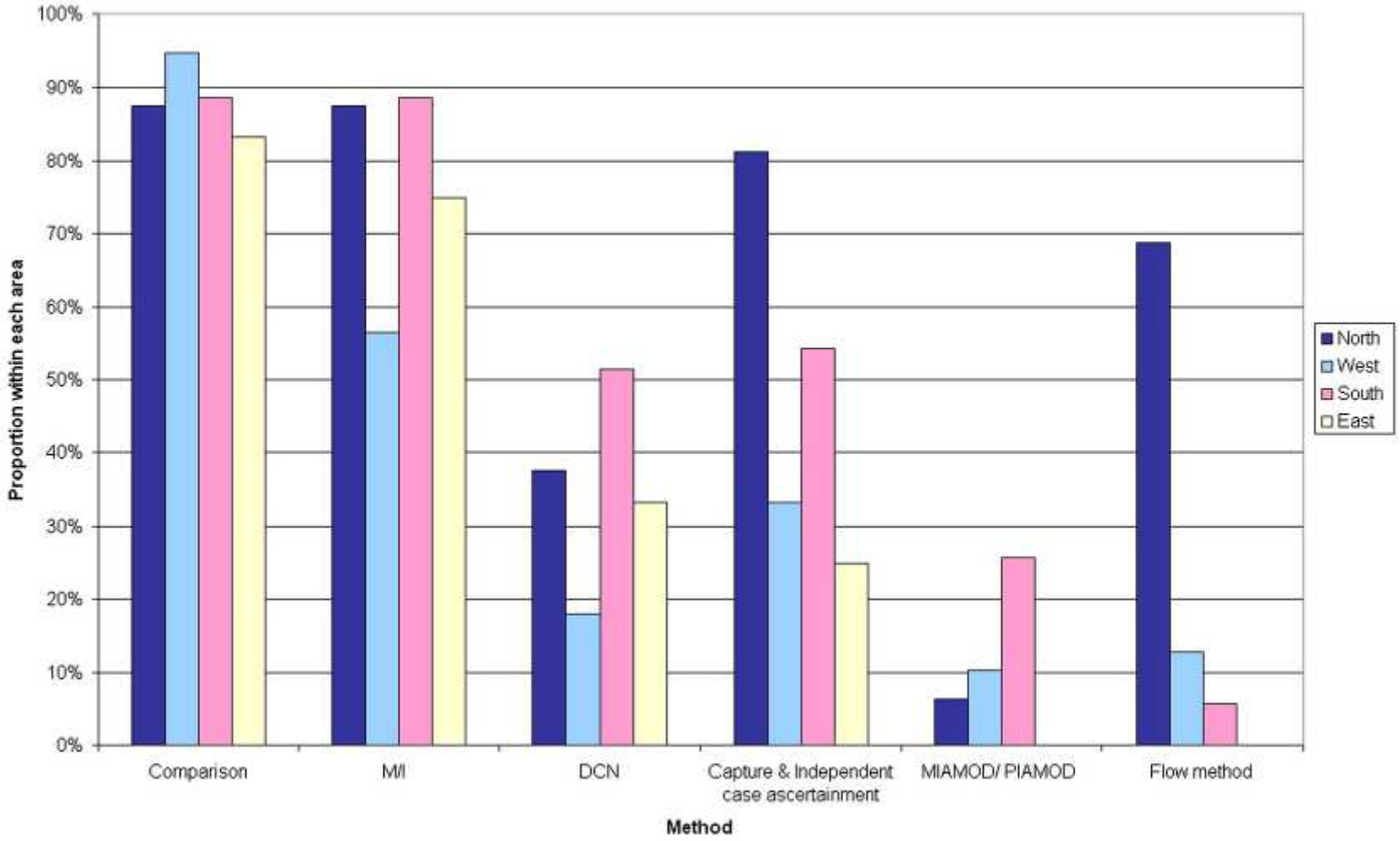
In the EU: 20 national and 82 regional CRs, covering 72% of the population

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

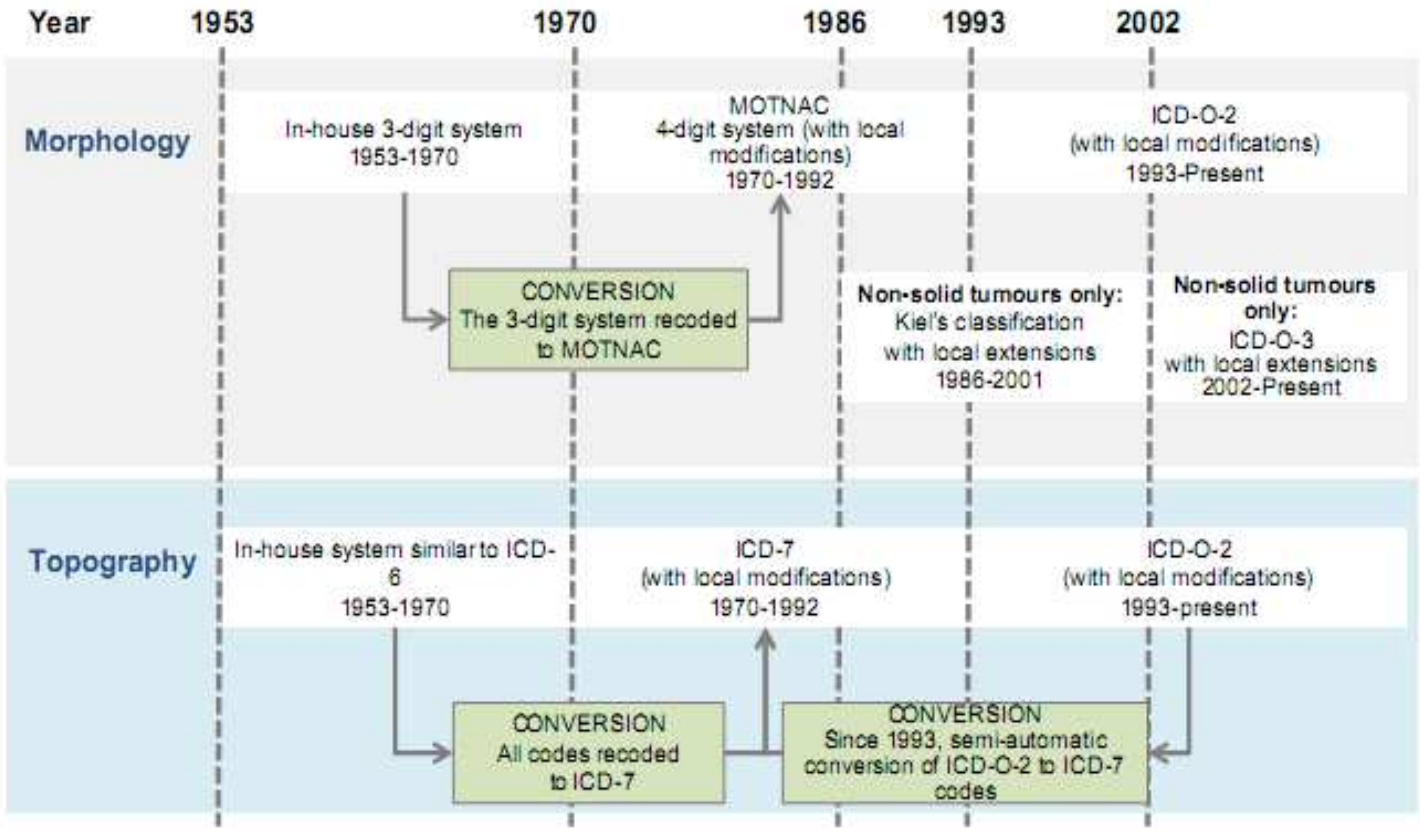


Fig. 1 – The standards of classification and coding of neoplasms followed, Norway 1953–2007.

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.

EVALUATION OF COMPARABILITY - EXAMPLE

- Topics:
  - The definition of incidence – what is the **incidence date**

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,[17] 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.[18] 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.

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

Cancer Incidence in Five Continents, Chapter 5

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

Age group (years)	Boys		Girls	
	Lowest	Highest	Lowest	Highest
0–4	< 13.7	> 25.6	< 11.3	> 23.2
5–9	< 8.9	> 16.5	< 7.0	> 12.7
10–14	< 9.2	> 16.3	< 7.9	> 14.9

Data quality at the Bulgarian National Cancer Registry: An overview of comparability, completeness, validity and timeliness

Nadya Dimitrova [a,*], Donald Maxwell Parkin [b]

Capture–recapture and DCN/M:I estimates of completeness of registration for the period 2006–2010, by cancer site and sex.

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

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

VALIDITY – EXAMPLE – CI5-X, 2003-2007

Males	MV%	DCO%	PSU%
Bulgaria	73,3	9,8	5,0
Croatia	77,0	5,0	3,9
Estonia	89,2	-	2,5
Latvia	71,5	11,2	2,4
Lithuania	83,8	4,9	2,5
Serbia	78,3	6,5	2,9
Slovakia	85,6	9,0	2,3
Slovenia	91,9	0,9	2,4
average	**81,3**	**6,8**	**3,0**

Tumori, 100: 399-405, 2014

Recent trends in cancer incidence: impact of risk factors, diagnostic activities and data quality of registration

Silvia Dehler[1], Simeon Tonev[2], Dimitri Korol[1], Sabine Rohrmann[3], and Nadya Dimitrova[2]

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

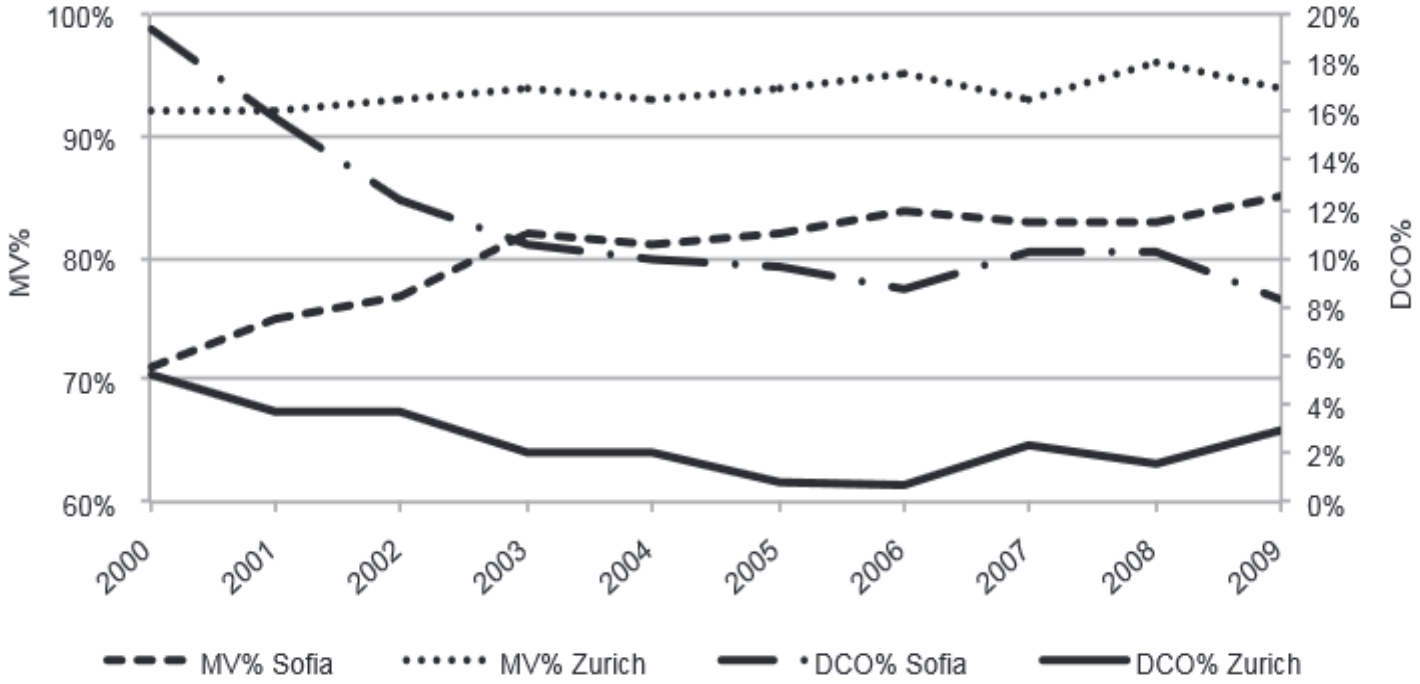


Figure 2 - Annual percentage of death-certificate-only (DCO) and morphologically verified (MV) cases in the years 2000-2009 in Sofia and the Canton of Zurich.

TIMELINESS

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

Table 4 – Registered cancer cases in Norway, 2005 as obtained from the registry extracted 29th November 2006 and 19th November 2007.

ICD-10	Site	Cases 2005		Difference	%
		2006	2007		
C00–96	All sites	24,199	24,730	531	2.2
C00	Lip	77	85	8	10.4
C01–02	Tongue	77	76	–1	–1.3
C03–06	Mouth, other	81	81	0	0.0
C07–08	Salivary glands	48	51	3	6.3
C09–14	Pharynx	132	127	–5	–3.8
C15	Oesophagus	198	197	–1	–0.5

TIMELINESS

- EUROCOURSE survey

Latency for completing 1 year of case ascertainment and releasing data of European cancer registries (in months).

	n	Mean	Median	Minimum	Maximum
Complete 1 year of incidence	113	21	18	4	60
Publish printed report	92	7	6	1	42
Publish data on internet	89	6	3	1	30

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

- Parkin DM, Chen VW, Ferlay J, Galceran J, Storm HH and Whelan SL. Comparability and Quality Control in Cancer Registration. IARC Technical Report №19, 1994. Lyon, International Agency for Research on Cancer, WHO and International Association of Cancer Registries
- Parkin DM, Plummer M. Comparability and quality of data. In: Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB (eds). Cancer Incidence in Five Continents, vol. VIII. Lyon: International Agency for Research on Cancer, 2002
- Parkin M, Bray F. Evaluation of data quality in the cancer registry: Principles and methods. Part II. Completeness. EJC, 45:756-764, 2009
- Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. Eur J Cancer. 2009 Mar; 45(5):747-55
- Larsen IK, Smastuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, Moller B. Data quality at the Cancer Registry of Norway: An overview of comparability, completeness, validity and timeliness. EJC, 45:1218-1231, 2009


I am an imperfect person
Loved by a perfect God