Cancer Epidemiology 39 (2015) 405–413

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

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

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

[a] Bulgarian National Cancer Registry, National Hospital of Oncology, 6 Plovdivsko pole str., 1756 Sofia, Bulgaria
[b] Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Charterhouse Square, London EC1M 6BQ, UK

ARTICLE INFO

Article history:
Received 11 December 2014
Received in revised form 24 March 2015
Accepted 25 March 2015
Available online 22 April 2015

Keywords:
Cancer registry
Data quality
Neoplasms
Comparability
Completeness
Validity
Timeliness

ABSTRACT

Reporting of neoplasms in Bulgaria has been compulsory since a directive from the Ministry of Health in 1951. The quality of cancer registry data has been estimated rather infrequently in past years.

We aimed to provide a comprehensive evaluation of the quality of the data at the Bulgarian National Cancer Registry (BNCR).

Quantitative and semi-quantitative methods were applied for cancers diagnosed during the whole period 1993–2010, and also for cases diagnosed in 2006–2010. The methods used include historic data methods, mortality-to-incidence ratios (M:I), capture–recapture and death-certificate methods, proportions of morphologically verified cases (MV%), death-certificate-only cases (DCO%), and cases with missing information (primary site unknown, PSU%; stage unknown, SU%).

The BNCR coding and classification systems follow international standards. The overall completeness was estimated at 92.6–94.7% for the period 2006–2010, with variations between cancer sites (86.7–98.5%). During the period 1993–2010, M:I decreased to 0.5 for males and 0.4 for females, MV increased to 87.4%, DCO and SU decreased to 4.8% and 18.8%, respectively, and PSU remained at the same level of about 4% for both sexes together. Sub-analysis revealed differences by site, sex and age groups. The comparison with other registries from the region showed similar incidence rates and directions of trends: M:I, MV% and DCO% that were not significantly different. The underreporting in 2008 and 2009 due to timely publication was estimated at an overall 0.8% and 0.5%, respectively.

The present review showed that the BNCR yields internationally comparable data that are reasonably accurate, timely, and close to complete, especially in recent years. This is a prerequisite for the BNCR to expand its role to more areas of cancer control.

1. Introduction

The Bulgarian National Cancer Registry (BNCR) was established in 1952. Reporting of neoplasms has been compulsory since a directive from the Ministry of Health in 1951. Nowadays, there are 13 regional cancer registries (RCRs) situated in the major cities across Bulgaria. They are part of the regional oncological centers. The National Cancer Registry is part of the National Oncological Hospital in the capital Sofia. Cancer registrars from RCRs collect information, using active and passive combined approaches, from all available data sources in their region; information is extracted and recorded in the regional database. All 13 regional databases are combined to create the national database.

The information about the patient, the tumor, and treatment is structured in more than 85 items in the database.

The quality of data at THE BNCR has been assessed rather infrequently in past years, historic data methods mainly being used. Some indicators for validity were published in volumes IX and X of cancer incidence in five continents (CI5) [1,2]. The completeness for registrations in 2001–2005 was estimated with the capture–recapture method at 94.7% [3]. The proportions of morphologically verified (MV) and death-certificate-only (DCO) cases have been regularly published in the annual reports since 2007 [4–9].

Considering their important role in cancer control, all modern cancer registries – even those operating in low- and middle-income settings – should be able to give some objective and preferably quantitative estimation of the quality of data that they have collected [10]. International organizations such as the European Network of Cancer Registries (ENCR, www.encr.eu) and the International Association of Cancer Registries (IACR,

* Corresponding author. Tel.: +359 2 8076367; mobile: +359 889 378720.
E-mail addresses: dimitrova.nadia@gmail.com (N. Dimitrova), d.m.parkin@qmul.ac.uk (D.M. Parkin).

http://dx.doi.org/10.1016/j.canep.2015.03.015
1877-7821/

www. iacr.com.fr) encourage this activity through relevant recommendations, published on their websites, and also through promoting the good practices among registries [11].

The aim of the present study is to provide a comprehensive evaluation of the quality of the most recent data at the BNCR in terms of comparability, completeness, validity and timeliness – the four main dimensions of data quality [10,12–14].

2. Materials and methods

Data for all cancer cases diagnosed in 1993–2010, as of 1st March 2012, were extracted from the BNCR. The whole time span was used for evaluation of comparability, completeness and validity, with special attention to the most recent 5-year period 2006–2010, using the methods described by Bray and Parkin [12,13]. The population and mortality figures were provided by the National Statistical Institute (www.nsi.bg). The vital status of the patients was updated as of 31st December 2011, according to the information from the civil registration system (ESGRAON).

Comparability is the extent to which classification and coding procedures applied at the BNCR accord with published international guidelines and standards [12]. The items covered here are topography, morphology, incidence date, basis of diagnosis, multiple primaries, and stage.

Completeness is defined as the extent to which all diagnosed neoplasms in Bulgaria are included in the registry database [13]. Semi-quantitative methods used here include the historic data method (stability of incidence rates over time, 1993–2010), comparison of incidence rates in different populations (2003–2007), shape of age-specific curves (2003–2007), incidence rates of childhood cancers (2005–2009), and the mortality-to-incidence ratio (the ratio 2005–2009 compared with 1 minus 5-year observed survival for cases diagnosed in 2001–2005). Capture–recapture and death-certificate (DCN/M:I) methods were used for quantitative estimation of completeness for the period 2006–2010 [13,14].

Validity refers to the accuracy of the recorded data at the BNCR and is defined as 'the proportion of cases in the registry with a given characteristic that truly have that attribute' [12]. The proportions of morphologically verified cases (MV%), death-certificate-only cases (DCO%), and cases with missing information (primary site unknown, PSU%; stage unknown, SU%) for the period 1993–2010 were analyzed. MV%, DCO% and SU% for the period 2006–2010 are evaluated by sex, age group and site. MV%, DCO% and PSU% for the period 2003–2007, as published in CI5-X [2], are compared with the corresponding estimates for the region, which allowed the identification of significant differences evaluated with statistical tests routinely used in CI5 series and described in detail elsewhere [15].

Timeliness is evaluated in terms of the time from diagnosis to the reporting of incidence in the annual reports of the BNCR.

3. Results

The main sources of information for cancer registration in Bulgaria are notifications of new cancer cases, hospital discharge records, results from pathology and other diagnostic laboratories, and death certificates. The procedures followed are shown in Fig. 1. The information on cancer cases was stored only on paper until 1993, when the computer database was created, and since then the records are maintained in electronic format.

3.1. Comparability

Incident cases in Bulgaria comprise all malignant and in situ neoplasms. Topography was coded according to the International Classification of Diseases (ICD): ICD7 (until 1968), ICD8 (1969–1979), ICD9 (1980–2004) and ICD10 (2005 to the present). Morphology has been coded according to ICD9 (corresponding to codes from ICD for oncology: ICD-O-1) since 1991. In 2005 all topography and morphology codes (registrations 1993–2004) were converted to ICD10, which incorporates morphology codes from ICD-O-2.

The *incidence date* has been recorded according to the European Network of Cancer Registries (ENCR) recommendations [16] since 2008; the date of morphological examination has the highest priority. For cases diagnosed before 2008, there was an in-house rule that the incidence date was the earliest date mentioning cancer from the medical documents of the patient. For the period 2000–2007, 36.9% of the cases had a different date on applying the ENCR rules. For these cases, the median difference between the ENCR-defined and in-house incidence dates was 12 days.

The *basis of diagnosis* is coded according to the rules of the International Agency for Research on Cancer (IARC) and the International Association of Cancer Registries (IACR), reproduced in ICD-O [17,18]; only 'the most valid' basis is recorded for a given cancer.

The recording and reporting of *multiple primary* tumors follow the IARC guidelines with respect to the groups of topography codes considered as single sites, and groups of morphology codes considered as histologically 'different', regardless of time between dates of diagnoses [19]. In 2008, the BNCR adopted the rule that two tumors of different laterality, but of the same morphology, diagnosed in paired organs (e.g. breast) should be registered separately [20]. This resulted in a slight increase in the breast cancer incidence rate (age-standardized, world standard per 100,000) from about 53 for the period 2003–2007 to 55 in 2008–2010, and an increase in the proportion of multiple tumors in breast cancer patients from about 8–10%, with no significant effect on other sites.

The coding of *stage* at the BNCR accords with the international guidelines for TNM, where applicable [21].

3.2. Completeness

Incidence rates for the period 1993–2010 for selected sites by sex are presented in Fig. 2. The *annual trends* appear as relatively constant changes, with no major fluctuations in the pattern except the steep increase in lung cancer incidence in 1998 (also observed for other cancers to a lesser degree). This was probably due to the fact that in that year the BNCR started to receive information also from death certificates, provided by the CRS, which was included in the calculation of incidence.

Incidence rates for the period 2003–2007 for all sites in Bulgaria are similar to those of other countries in South-Eastern Europe (SEE), and also directions of trends for the period 1999–2008 are alike for most of the cancer sites, apart from lung and cervical cancers [2,22] (Fig. 3, Table 1).

The shapes of *age-specific curves* of incidence rates are typical for the major cancers, which implies no variability in completeness by age. Fig. 4 shows that for selected cancers in Bulgaria, age-specific curves are identical with those from other SEE countries [2].

The age-specific incidence rates for *childhood cancer* in the period 2005–2009 are shown by sex in Table 2. The values are within the expected limits (the upper and lower deciles for childhood cancer incidence rates published in volume VIII of CI5) for each age group [13].

The assessment of the mortality-to-incidence ratio (M:I) compared with 1 minus 5-year observed survival (*1 – survival*) showed that there was a strong coherence: e.g. cancers with poor survival rates (such as those of the pancreas, liver, esophagus,