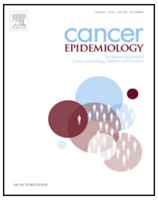
Cancer Epidemiology 43 (2016) 76–86

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 Singapore Cancer Registry: An overview of comparability, completeness, validity and timeliness

Janice Wing Mei Fung[a,*], Sandra Bee Lay Lim[a], Huili Zheng[a], William Ying Tat Ho[a], Bee Guat Lee[a], Khuan Yew Chow[a], Hin Peng Lee[b]

[a] *National Registry of Diseases Office, Singapore Cancer Registry, Health Promotion Board, 3 Second Hospital Avenue, Singapore 168937, Singapore*
[b] *Saw Swee Hock School of Public Health, National University of Singapore, 12 Science Drive 2 #10-01 Singapore 117549, Singapore*

ARTICLE INFO

Article history:
Received 4 March 2016
Received in revised form 21 June 2016
Accepted 23 June 2016
Available online 9 July 2016

Keywords:
Cancer registry
Data quality
Neoplasms
Comparability
Completeness
Validity
Accuracy
Timeliness

ABSTRACT

Aim: To provide a comprehensive evaluation of the quality of the data at the Singapore Cancer Registry (SCR).

Methods: Quantitative and semi-quantitative methods were used to assess the comparability, completeness, accuracy and timeliness of data for the period of 1968–2013, with focus on the period 2008–2012.

Results: The SCR coding and classification systems follow international standards. The overall completeness was estimated at 98.1% using the flow method and 97.5% using the capture-recapture method, for the period of 2008–2012. For the same period, 91.9% of the cases were morphologically verified (site-specific range: 40.4–100%) with 1.1% DCO cases. The under-reporting in 2011 and 2012 due to timely publication was estimated at 0.03% and 0.51% respectively.

Conclusion: This review shows that the processes in place at the SCR yields data which are internationally comparable, relatively complete, valid, and timely, allowing for greater confidence in the use of quality data in the areas of cancer prevention, treatment and control.

Contents

---

[*] Corresponding author.
E-mail addresses: janice_fung@hpb.gov.sg, wingmei.fung@gmail.com (J.W.M. Fung).

http://dx.doi.org/10.1016/j.canep.2016.06.006
1877-7821/

1. Introduction

Established in 1968, the Singapore Cancer Registry (SCR) is a population-based national registry with a repository of information on borderline, in-situ and malignant tumours and benign neoplasms which were diagnosed or treated in Singapore, with data collected by all available sources nation-wide [1]. Singapore is a multi-ethnic country with the Chinese, the Malays and the Indians being the three main ethnic groups in its resident population, which includes Singapore citizens and permanent residents. Residents comprised 94.5% of the total population in 1980, 81.2% in 2000, and 71.2% in 2013 due to a large influx of foreign workers in the past decades [2,3]. The SCR registers all cases of cancer diagnosed in Singapore, but for this study, the incidence rates pertain only to Singapore's resident population. The registry data is used for international and national benchmarking, guiding policy direction and in research [4]. The high quality of data is hence critical for accurate decision making [5,6].

The reporting of neoplasms has been mandatory since 2009, when the Ministry of Health enacted the National Registry of Diseases (NRD) Act [7] to provide statutory regulations on reporting. All hospitals, laboratories, and healthcare institutions in Singapore were required to report cancer cases [1,7,8]. Prior to this, cancer notifications were voluntary.

This study presents the first comprehensive evaluation of the data quality of SCR in line with the evaluation guidelines published by the International Agency for Research on Cancer (IARC) [9] and techniques covered by Bray and Parkin [10,11]. The quality of the data collected by the SCR will be assessed in the four indicative areas of comparability, completeness, validity and timeliness.

2. Material

Sources of notification

The sources of notification for the SCR had grown increasingly comprehensive since its establishment in 1968. Registrations were initially based on voluntary notifications received from clinicians [12]. From 2001 onwards, the pool of sources was enlarged through data obtained from pathology records, hospital discharge summaries and death certificates [2]. This was enhanced by the enactment of the NRD Act in 2009 which mandated medical professionals to report all histologically verified and clinically diagnosed cancer incidents within three months from the date of diagnosis, using a structured notification form [1,8].

Modes of notification include usage of hardcopy forms or via an online national healthcare portal [4], with monthly tracking done by registry staff. A secure electronic system also facilitates the information transfer of the Hospital Inpatients Discharge Summary (HIDS) listings. As a majority of cancer cases are diagnosed histologically, the SCR receives monthly pathology reports from all public and private laboratories. Haematological, radiotherapy, and nuclear medicine information from treatment centres are also sent to the SCR on a regular basis. The workflow of the notification process is illustrated in Fig. 1.

Patient data are collected using the unique national identity number issued to Singapore residents. For non-residents, their passport numbers or foreign identification numbers (FIN) are used [1]. These unique numbers minimize the risk of potential duplication of records in the SCR database amidst multiple notification sources.

Data are captured via electronic transmission from hospitals to the registry using the National Registry of Diseases System (NRDS) since 2004, and additional information is collected manually from hardcopy medical records. During the data collection process, verification of information for accuracy and completeness is performed manually by a team of nine registry staff and a visiting consultant pathologist [13]. Retrospectively, a separate team of staff perform annual audits on the data collected to ensure that the level of data accuracy is at least 95%.

Mortality data of all patients are obtained from the National Death Registry on a two-monthly basis via secured file transfer channels. The certification of death is virtually complete in Singapore [13].

3. Methods

3.1. Comparability

Comparability is the extent of which coding and collection practices adhere to international guidelines [14]. In this study, comparability is examined by reviewing the registry's standards, definition and practices in incidence date and basis of diagnosis, topography and morphology, coding of multiple primaries, and staging [15].

3.2. Completeness

Completeness is the extent to which all diagnosed cancer cases in Singapore are captured in the registry database [10,14–16]. This allows us to determine whether incidence rates and survival proportions reflect their true values. Taking reference from Parkin and Bray [10], we used the historic data approach (stability of incidence rates over time (1968–2013), shape of age-specific curves for selected cancers (2008–2012), and age-specific incidence rates of childhood cancer (2008–2012) compared with the reference deciles for childhood cancer published in the Cancer Incidence in Five Continents (CI5), Volume X [9]). Segi's world population was used for direct standardisation in calculating age-standardised rates. Other semi-quantitative methods used include the mortality/incidence (M:I) ratio (2008–2012) versus one minus five years relative survival (2003–2012), and the number of notifications per case (2008–2012) [10,15,17].

Quantitative methods included the capture-recapture method (2008–2012) and the flow method for cases diagnosed in 2008 and followed up until 31 December, 2012. The capture-recapture method assumes that all the notification sources are independent of each other and all patients have the same probability of being captured in the Registry. However, these assumptions are likely to be violated in cancer registration. The notification sources can be broadly classified into: clinical sources, pathological sources and death certificates. When substantially two of three groups are dependent on each other, it is possible to correct for the dependency between the sources by pooling the two groups into one broad group and compare it with the remainder group by a two-way capture-recapture method [18,19].

The flow method models the flow of individuals through the case ascertainment process from diagnosis to registration, taking into account of the time since diagnosis [20]. Two datasets were used to run the flow method: cases diagnosed in 2008 and cancer cases who died in 2012. Using the flow method, we computed the probability of patient being registered, missing or lost, and we plotted the probability of patient being registered over the years since diagnosis.

All statistical analysis were performed using Stata SE Version 13.

3.3. Validity

Validity refers to the accuracy of the data, and is defined by Bray & Parkin as *the proportion of cases in the registry with a given characteristic which truly have this attribute* [11]. The proportions of