ACADEMIC HOSPITAL JOURNAL

Volume 5, No. 1, March 2023

ISSN 2615-6253 (ONLINE)
ISSN 2615-6245 (PRINT)

Evaluation of Completeness of Cancer Data Variables on Data Quality in Hospital-Based Cancer Registration Activities at Dharmais Cancer Hospital

***Grace Shalmont[1], Pradnya Sri Rahayu[1], Susanna Hilda Hutajulu[2], and Lutfan Lazuardi[2]**

[1]National Cancer Centre, Dharmais Cancer Hospital, Jakarta
[2]Universitas Gadjah Mada, Yogyakarta
*Corresponding author: graceshalmont@yahoo.com

Submitted: December 2022 Reviewed: February 2023 Publish: March 2023

Abstract

Background: Cancer is a major burden of disease worldwide, including in Indonesia. As an effort to control the burden of cancer, WHO established National Cancer Control Programs (NCCP) where cancer registration is one of the key points. Dharmais Cancer Center Hospital, appointed as the national cancer center, has the responsibility to conduct a cancer registry in Indonesia. The good quality data of cancer registry according to international standards is beneficial to describe the cancer burden in the country. In Dharmais Cancer Center Hospital, microscopic verification is one of the variables that has not been qualified. Therefore, it is important to evaluate the completeness of cancer data variables toward data cancer quality on hospital-based cancer registry of Dharmais Cancer Center Hospital. To assess the quality of cancer data based on microscopic verification, to evaluate the completeness of hospital-based cancer registry variables and the quality of data based on microscopic verification between complete and incomplete variable groups.

Materials and Methods: This quantitative research is an observational study (non-experimental) with cross-sectional study design. It utilizes secondary data from hospital-based cancer registry of Dharmais Cancer Center Hospital for incidence year 2013-2017.

Results: Data quality of microscopic verification that assessed on a complete data group is 87,8% and for overall cancer cases is 62%. Among social variables, identity numbers are the most incomplete variable, which is 39%. While among tumor data variables, stage is also the most incomplete variable with 82% data. There are differences between the quality of data based on microscopic verification with the completeness of data, especially among social data variables and tumor data variables.

Conclusion: The quality data based on microscopic verification that is assessed on a complete variable group is better than microscopic verification on overall cancer cases. The incomplete variables among social variables are identity number, date of birth, address, and district/province. Whereas on tumor variables, the incomplete variables are stage, treatment, metastasis, and laterality. The completeness of cancer data has an important role on data quality based on microscopic verification mainly on social and tumor variables. Improvement and strengthening particularly on management and technical aspects of cancer registration are indispensable.

Keywords: Cancer data variables, hospital-based cancer registry, microscopic, validity, verification

1. Introduction

Based on the background, the researcher intends to research differences in clinical outcomes based on nutritional status in acute ischemic stroke patients at Academic Hospital UGM. Cancer is a non-communicable disease that is currently a major burden of disease worldwide, including in Indonesia. According to Globocan 2020 there is an increase in the number of new cases in the world by 19,292,789 cases and the number of deaths by 9,958,133 deaths. Data estimated for Indonesia, there were an additional 396,914 new cases and 234,511 deaths, with a total population of 273,523,621. Morbidity and mortality from non-communicable diseases are expected to continue to increase, and in particular cancer is estimated to have an additional 20,000,000 new cases by 2025, and about 4/5 cases are found in lower-middle-income countries (1).

In an effort to control the burden of cancer, the World Health Organization established National Cancer Control Programmes (NCCP) which include preventive programs, early detection and screening, diagnosis and therapy, palliative care, medical rehabilitation, cancer surveillance and registration, and research (2). The Ministry of Health of Indonesia has assigned a national action plan in cancer control and cancer registration is the essential part of the program, which has an important role to provide the good quality cancer data so that it can be used as a basis for making a cancer control program policy (3). Dharmais Cancer Center Hospital has been appointed by the ministry of health as national cancer burden data center to conduct hospital-based cancer registry and population-based cancer registry (4).

Hospital-based cancer registry is the main source of population-based cancer registry. It also can be used as a review of health services in the hospital, hospital administrative purposes, and a source of data for research. In this activity, all cancer data within the hospital is recorded. In order to complete the microscopic verification variables, pathology-based cancer registry is the main data source (5,6).

At the moment, Indonesia cancer data cannot yet be included in CI5 vol X and XI because the quality of cancer data according to the standard has not yet been met. The last cancer data collection for CI5 volume XII is a cancer data report for 2013-2017, which has been collected in November 2021, and is currently waiting for assessment from WHO (7).

The quality of data from cancer registries is important in describing the extent of the cancer burden, a source for etiologic studies, and for monitoring and assessing cancer prevention and control activities (8).

Completeness, comparability, validity, and timeliness are four aspects in the evaluation of cancer registry data quality. Data collection and completeness of data variables from each data source is one of the important things that affect cancer data quality. Yang et al. in their study suggested that there was a difference in overall survival between data groups with complete and incomplete variables. Yang et al. also suggested that it is important to improve the documentation and quality of cancer data for optimal data utilization. A good data source with complete cancer data variables is expected to produce good output, not only in quantity, but of course in terms of data quality (9).

Considering the quality of cancer data in the hospital, particularly in the aspect of validity, microscopic verification has not been able to meet the standards, it is necessary to conduct research from the managerial aspect starting from hospital-based cancer registration in Dharmais Cancer Center Hospital before it can be developed and carried out on population-based cancer registration data.

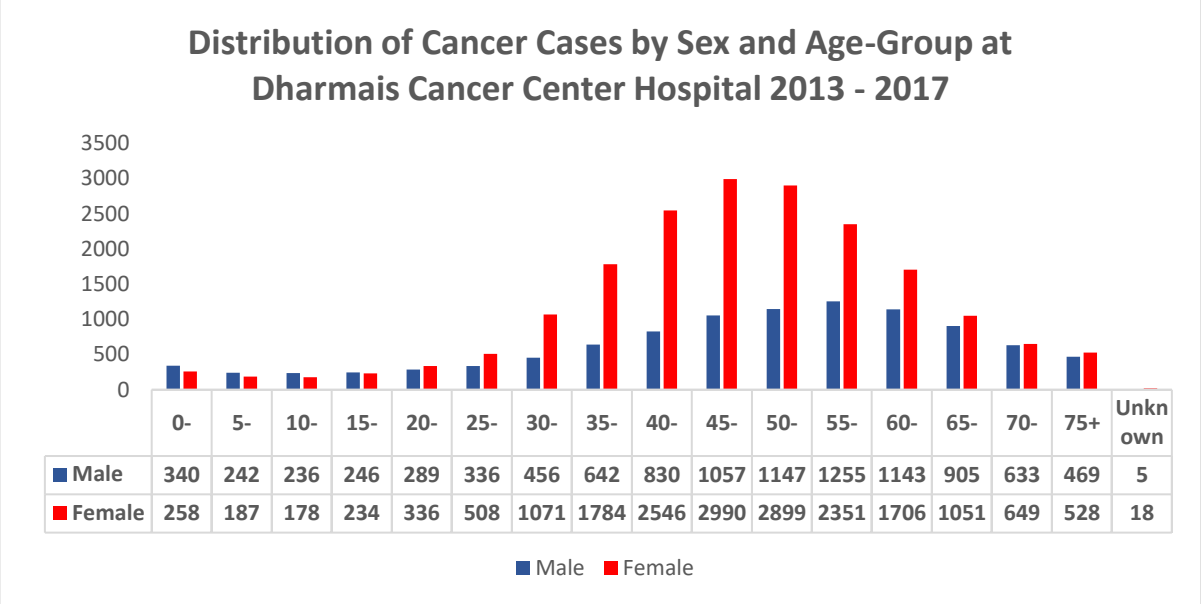
2. Materials and Method

This quantitative research is an observational study (non-experimental) with cross-sectional study design. It utilizes secondary data from hospital-based cancer registry of Dharmais Cancer Center Hospital for incidence year in 2013-2017.

3. Results

Hospital-based cancer registry of Dharmais Cancer Center Hospital has recorded 29525 cases for incidence year 2013-2017. 4406 cases in 2013;

6519 in 2014; 6400 cases in 2015; 6115 cases in 2016; and 6085 cases in 2017. All cases were included in the analysis. The ratio of male and female is 3:7 (34,6% for male and 65,4% for female). The distribution of cancer cases for male reached its peak in the age-group of 45-64 years (45% of all cases), while for female the distribution of cancer cases peaked in the younger age-group, 40-59 years (56% of all cases). In children (0-14 years) the number of cases for boys is greater than girls, around 200 cases. However, when compared to the total number of cases, cancer cases in boys is 8% of total cases while in girls it is 3% of total cases.



	0-	5-	10-	15-	20-	25-	30-	35-	40-	45-	50-	55-	60-	65-	70-	75+	Unknown
Male	340	242	236	246	289	336	456	642	830	1057	1147	1255	1143	905	633	469	5
Female	258	187	178	234	336	508	1071	1784	2546	2990	2899	2351	1706	1051	649	528	18

Figure 1. Distribution of cancer cases by sex and age-group at Dharmais Cancer Center Hospital in 2013 - 2017

Top 3 most common cancer of all cases are breast, cervix, and lung cancer. For male are lung, nasopharyngeal, and non-hodgkin lymphoma. For female are breast, cervix, and ovary cancer.

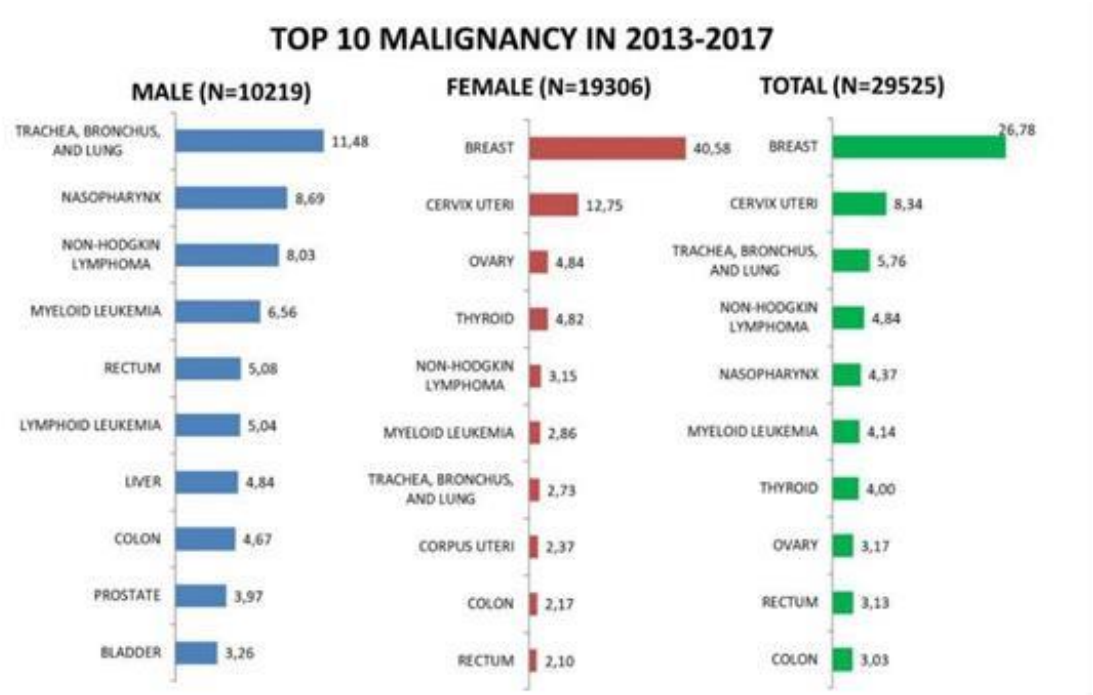


Figure 2. Top 10 malignancy by sex of all cancer cases at Dharmais Cancer Center Hospital 2013-2017

Seventeen independent and one dependent variables were included in the analysis. It was basic variables that are available on cancer registry form. The independent variables contain 3 main group variables (social, tumor, and follow-up). Identity number, date of birth, sex, address, and district/province are variables for social data. Age, incidence date, topography, morphology, tumor behavior, basic diagnosis, grade, stage, treatment, metastasis, and laterality are variables for tumor data. Then for follow-up data there are dates of last contact and last contact status. The dependent variable is the validity variable, which is microscopic verification.

Of all 5 years (2013-2017) cancer cases, 18174 (62%) cases were microscopically verified. If cases with incomplete variables are excluded, then the number of cases with complete variables is 2052 (7%) cases and cases that are microscopically verified are 1802 (87,8%) cases. Afterward, for cases that have at least 1 incomplete variable is 27473 (93%) cases. Cases with incomplete data variables are 499 (1,69%) cases and those microscopically verified are 135 (27,05%) cases.

Table 1. Comparison of microscopic verification between cancer cases group based on variable completeness

Categories	N (%)	Microscopic Verification N (%)
All Cases	29525 (100.0)	18174 (62.0)
Cases with all complete variables	2052 (7.0)	1802 (87.8)
Cases with all incomplete variables	499 (1.69)	135 (27.05)

The social data, such as identity number, date of birth, sex, address, and district/province, shows there is incomplete data. Identity number completeness is 61%, date of birth is 99,9%, sex is 99,9%, address is 96%, and district/province is 96%. For tumor data, incidence data, topography, morphology, basic diagnosis, and grade shows completeness to 100%. The completeness for age is 99%, stage is 18%, treatment is 77%, metastasis is 73%, and laterality is 86%. While for follow-up data, all variables show completeness to 100%.

Table 2. Distribution of completeness of cancer data variables

Variable		N (%)
Total Cases		29525 (100.0)
	Microscopic Verification	
	Microscopically Verified	18174 (62.0)
	Non-Microscopically Verified	11351 (38.0)
Social Data	**Identity Number**	
	Complete	18117 (61.0)
	Incomplete	11408 (39.0)
	Date of Birth	
	Complete	29502 (99.99)
	Incomplete	23 (0.01)
	Sex	
	Complete	29525 (100.0)
	Incomplete	0 (0)
	Address	
	Complete	28226 (96.0)
	Incomplete	1299 (4.0)
	District/Province	
	Complete	28226 (96.0)
	Incomplete	1299 (4.0)
Tumor Data	**Age**	
	Complete	29502 (99.99)
	Incomplete	23 (0.01)
	Incidence Date	
	Complete	29525 (100.0)
	Incomplete	0 (0)
	Topography	
	Complete	29525 (100.0)
	Incomplete	0 (0)
	Morphology	
	Complete	29525 (100.0)
	Incomplete	0 (0)
	Basic of Diagnosis	
	Complete	29525 (100.0)
	Incomplete	0 (0)
	Grade	
	Complete	29525 (100.0)
	Incomplete	0 (0)
	Stage	
	Complete	5343 (18.0)
	Incomplete	24182 (82.0)
	Treatment	
	Complete	22693 (77.0)
	Incomplete	6832 (23.0)
	Metastasis	
	Complete	21537 (73.0)
	Incomplete	7988 (27.0)
	Laterality	
	Complete	25418 (86.0)
	Incomplete	4107 (14.0)

Follow-Up Data	**Date of Last Contact**	
	Complete	29525 (100.0)
	Incomplete	0 (0)
	Last Contact Status	
	Complete	29525 (100.0)
	Incomplete	0 (0)

Of all cases, there are 18174 (62%) microscopically verified and 11351 (38%) not microscopically verified. Table 6 concludes that variables from data social group have the most incomplete data, 4 out of 5 variables.

Table 3. Bivariate analysis of cancer data variable completeness with microscopic data quality

Variable	Microscopically Verified N (%)	Non-Microscopically Verified N (%)	*p-value*
Identity Number			
Complete	11483 (63.0)	6634 (58.0)	0.000
Incomplete	6691 (37.0)	4717 (42.0)	
Date of Birth			
Complete	18152 (99.99)	11350 (100.0)	0.001
Incomplete	22 (0.01)	1 (0)	
Sex			
Complete	18174 (100.0)	11351 (100.0)	0.206
Incomplete	0 (0)	1 (0)	
Address			
Complete	17002 (94.0)	11224 (99.0)	0.000
Incomplete	1172 (6.0)	127 (1.0)	
District/Province			
Complete	17002 (94.0)	11224 (99.0)	0.000
Incomplete	1172 (6.0)	127 (1.0)	
Age			
Complete	18152 (99.99)	11350 (100.0)	0.001
Incomplete	22 (0.01)	1 (0)	
Incidence Date			
Complete	18174 (100.0)	11351 (100.0)	N/A
Incomplete	0 (0)	0 (0)	
Topography			
Complete	18174 (100.0)	11351 (100.0)	N/A
Incomplete	0 (0)	0 (0)	
Morphology			
Complete	18174 (100.0)	11351 (100.0)	N/A
Incomplete	0 (0)	0 (0)	
Basic of Diagnosis			
Complete	18174 (100.0)	11351 (100.0)	N/A
Incomplete	0 (0)	0 (0)	
Grade			
Complete	18174 (100.0)	11351 (100.0)	N/A
Incomplete	0 (0)	0 (0)	
Stage			
Complete	18174 (100.0)	11351 (100.0)	N/A
Incomplete	0 (0)	0 (0)	
Treatment			
Complete	18174 (100.0)	11351 (100.0)	N/A
Incomplete	0 (0)	0 (0)	
Metastasis			
Complete	18174 (100.0)	11351 (100.0)	N/A
Incomplete	0 (0)	0 (0)	
Laterality			
Complete	3977 (22.0)	1366 (12.0)	0.000
Incomplete	14197 (78.0)	9985 (88.0)	
Date of Last Contact			
Complete	15094 (83.0)	7599 (67.0)	N/A
Incomplete	3080 (17.0)	3752 (33.0)	

Last Contact Status			
Complete	13856 (76.0)	7681 (68.0)	N/A
Incomplete	4318 (24.0)	3670 (32.0)	

Bivariate analysis between completeness of cancer data toward data quality that assessed from microscopically verified showed that there are differences in data quality especially in the variables of identity number, date of birth, address, district/province, age, stage, treatment, metastasis, and laterality with $p<0.05$.

4. Discussion

The analysis showed that of all cancer cases in 2013-2017 from the hospital-based cancer registry at Dharmais Cancer Center Hospital, the percentage of microscopically verified cases was 62%. By world standards, the indicator of the percentage of microscopically verified cases is still below the WHO standard of 75%. Hospital-based cancer registration data at RST Regional Cancer Hospital, Nagpur, India, showed microscopic verification of cancer cases in 2013-2014 was 88%. Hospital-based cancer registry data at Malabar Cancer Center, Kerala, India, showed microscopic verification of cancer cases in 2012-2014, amounting to 86% in men and 91% in women (10). Data from the Bangkok Cancer Registry from 9 government cancer hospitals showed microscopic verification of cancer cases in 2010-2012 was 62.6% in men and 74.2% in women (11).

Calculating the validity of microscopic verification when derived from cases with complete data variables, the percentage of microscopically verified cases increased to 87.8%. Consistent with data from a hospital-based cancer registry in Luanda, Angola, with incidence year in 2012-2016. Of the cancer cases with complete data variables, the percentage of microscopically verified cases of 92.3% met the standards set by WHO (12). However, if the calculation of microscopic verification comes from cases with incomplete data variables, it decreases significantly to 27.05. The number of cancer cases with at least 1 incomplete cancer data variable was 27473 (93%) cases.

From research conducted by Plichta et al., using breast cancer data derived from the National Cancer Data Base (NCDB) and SEER, breast cancer cases with at least 1 incomplete cancer data variable were 29% and 13% (13). Based on this study, incomplete variables from social data were found including variables of identity number, date of birth, sex, address, district/province, and tumor data variables including age, stage, treatment, metastasis, and laterality. Tumor data are age, stage, treatment, metastasis, and laterality. Research by Plichta et al. showed that the most incomplete variables came from tumor data (13,14).

Research conducted by Sirirungreung et al. with hospital-based cancer registration data in 2012-2014 taken from the National Cancer Institute Thailand, showed that social data like provinces and from tumor data such as grade, laterality, and morphology were the highest incomplete data variables compared to other variables. External factors that influence the incompleteness of social data variables are the possibility that patients do not bring an Identity Card or Family Card at the time of registration so that social data information becomes incomplete (8).

Some external factors that cause incomplete data on tumor variables are the absence of regulations that require patients to bring a complete resume from the previous referring hospital, resulting in a lack of information about initial stage data and treatment that has been carried out. Other factors such as the results of anatomical and radiological pathology expertise from the referring hospital were not carried or lost by the patient, so that information about tumor size, lymph node involvement, metastasis and laterality was incomplete. Until now, the incompleteness of data variables doesn't affect to national health insurance claims process.

An internal factor that can affect the completeness of social variable data is that the admission officer does not complete the patient's social data variables at the time of registration.

Meanwhile, internal factor that can affect the completeness of tumor variable data was the results of anatomical and radiological pathology expertise from the referring hospital brought by the patient, not stored in the medical record by doctors or nurses, so that information regarding tumor size, lymph node involvement, metastasis and laterality was incomplete. Other things such as the compliance of the doctor in charge of the patient in filling out the medical record completely, such as the absence of stage data recorded in the patient's medical record that can be extracted by the registrar. In addition, another internal factor that may cause incomplete data is the understanding and thoroughness of registrar staff in looking for these variables in medical records.

Elbasmi et al, in their study comparing breast cancer data extracted by registrars and oncologists, found that there were significant differences between the extracted data, especially the variables of morphology, grade, stage, and treatment (15). The lack of use of electronic medical records may be one of the factors that play a role in the completeness of data variables (16). In 2013-2017, medical records at Dharmais Cancer Center Hospital used conventional medical records and did not use electronic medical records. The absence of the use of electronic medical records can be one of the factors that play a role in the completeness of data variables. The existence of electronic medical records is expected to facilitate registrars in abstraction of cancer data (16). In this case, further research needs to be done to assess the relationship between the use of electronic medical records and the completeness of cancer data variables.

Bivariate analysis between data quality assessment based on microscopic verification between complete variables and incomplete variables, showed there were differences in data quality especially in the variables of identity number, date of birth, address, district/province, age, stage, treatment, metastasis, and laterality with $p<0.05$. From the literature, no one has directly linked the completeness of cancer data variables with microscopic verification.

Yang et al. in their study suggested that there was a difference in overall survival between groups with complete and incomplete data variables (9). Overall survival in patients with lung cancer, breast cancer, and prostate cancer is better in groups with complete data variables than in groups with incomplete data variables. Consistent with Plichta et al. also revealed the same thing, that overall survival in groups with complete data variables was better than groups with incomplete variables. A good data source accompanied by the completeness of cancer data variables is expected to produce good output, not only in quantity, but of course in terms of data quality (9). The need for access to good data sources and complete cancer registration data variables, so that the data output from cancer registration does not cause bias and can be optimally utilized (17,18).

The limitation of this study is that it uses the minimum variables available in cancer registration activities. There may be differences in results if there are additional variables studied. Another limitation is that this study aims to describe the quality of microscopic verification data based on the completeness of hospital-based cancer registration data variables. Further research needs to be done such as studies to find factors that play a role in the completeness of cancer data variables, studies that use data coverage larger than population-based cancer registrations, studies to see the relationship between the completeness of cancer data variables with survival, and others.

5. Conclusion

This research concludes that the data qualified by microscopic verification from complete cancer data group better than microscopic verification of overall cases. Social data group has the most incomplete data compared to tumor and follow-up data group, with identity number as the highest incomplete variable. There are also differences of data quality that are microscopically verified between complete and incomplete cancer data group, particularly on identity number, date of birth, address, district/province, age, stage, treatment, metastasis, and laterality.

The key point of the successed cancer registry is commitment from various parties within the hospital, from director, health workers, and non-health workers, involved to support cancer registration. Other than that, the strengthening of managerial aspect and technical aspect is a must. Human sources, capacity building, facilities, monitoring and evaluation, and improvements of the record system are things that must be considered. One final point, enhancement of communication with referring hospitals to include patient resumes, such as examination expertise results and treatment history, thereby patient information for cancer registry can be more complete.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209–49.
2. World Health Organization. National Cancer Control Programmes [Internet]. 2022. Available from: https://www.who.int/publications-detail-redirect/national-cancer-control-programmes
3. 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.
4. Kementerian Kesehatan RI. Pedoman Sistem Registrasi Kanker di Indonesia, Jakarta. Kementerian Kesehatan RI; 2018.
5. Mohammadzadeh Z, Ghazisaeedi M, Nahvijou A, Kalhori SRN, Davoodi S, Zendehdel K. Systematic Review of Hospital Based Cancer Registries (HBCRs): Necessary Tool to Improve Quality of Care in Cancer Patients. Asian Pac J Cancer Prev. 2017;18(8):2027–33.
6. Bernal-Delgado EE, Martos C, Martínez N, Chirlaque MD, Márquez M, Navarro C, et al. Is hospital discharge administrative data an appropriate source of information for cancer registries purposes? Some insights from four Spanish registries. BMC Health Serv Res. 2010 Jan 8;10:9.
7. National Cancer Institute. Cancer Incidence Rates Adjusted for Reporting Delay [Internet]. 2022 [cited 2022 Jan 10]. Available from: https://surveillance.cancer.gov/delay/
8. Sirirungreung A, Buasom R, Jiraphongsa C, Sangrajrang S. Data Reliability and Coding Completeness of Cancer Registry Information Using Reabstracting Method in the National Cancer Institute: Thailand, 2012 to 2014. J Glob Oncol. 2018 Sep;4:1–9.
9. Yang DX, Khera R, Miccio JA, Jairam V, Chang E, Yu JB, et al. Prevalence of Missing Data in the National Cancer Database and Association With Overall Survival. JAMA Netw Open. 2021 Mar 23;4(3):e211793.
10. Swaminathan S, Katoch V. Consolidated Report of Hospital Based Cancer Registries 2012-2014. Indian Council of Medical Research; 2016.
11. Insamran W, Sangrajrang S. National Cancer Control Program of Thailand. Asian Pac J Cancer Prev. 2020 Mar;21(3):577–82.
12. Miguel F, Bento MJ, de Lacerda GF, Weiderpass E, Santos LL. A hospital-based cancer registry in Luanda, Angola: the Instituto Angolano de Controlo do Cancer (IACC) Cancer registry. Infect Agent Cancer. 2019;14:35.
13. Plichta JK, Rushing CN, Lewis HC, Blazer DG, Hyslop T, Greenup RA. Missing data in breast cancer: Relationship with survival in national databases. JCO. 2020 May 20;38(15_suppl):e19114–e19114.
14. Eisemann N, Waldmann A, Katalinic A. Imputation of missing values of tumour stage in population-based cancer registration. BMC Med Res Methodol. 2011 Sep 19;11:129.
15. Elbasmi AA, Fayaz MS, Al-Mohanadi S, Al-Nesf Y, Al-Awadi A. Reliability of the Kuwait Cancer Registry: a comparison between breast cancer data collected by clinical oncologists and registry staff. Asian Pac J Cancer Prev. 2010 Jan 1;11(3):735–8.
16. Pollock AM, Vickers N. Reducing DCO registrations through electronic matching of cancer registry data and routine hospital data. Br J Cancer. 2000 Feb;82(3):712–7.
17. Wanner M, Matthes KL, Korol D, Dehler S, Rohrmann S. Indicators of Data Quality at the Cancer Registry Zurich and Zug in Switzerland. Biomed Res Int. 2018;2018:7656197.
18. Brackley ME, Penning MJ, Lesperance ML. In the absence of cancer registry data, is it sensible to assess incidence using hospital separation records? Int J Equity Health. 2006 Oct 6;5:12.