

# Data Reliability and Coding Completeness of Cancer Registry Information Using Reabstracting Method in the National Cancer Institute: Thailand, 2012 to 2014

**Purpose** Data quality is a core value of cancer registries, which bring about greater understanding of cancer distribution and determinants. Thailand established its cancer registry in 1986; however, studies focusing on data reliability have been limited. This study aimed to assess the coding completeness and reliability of the National Cancer Institute (NCI) hospital-based cancer registry, Thailand.

**Methods** This study was conducted using the reabstracting method. We focused on seven cancer sites—the colon, rectum, liver, lung, breast, cervix, and prostate—registered between 2012 and 2014 in the NCI hospital-based cancer registry. Missing data were identified for coding completeness calculation among important variables. The agreement rate and  $\kappa$  coefficient were computed to represent data reliability.

**Results** For reabstracting, we retrieved 957 medical records from a total of 5,462. These were selected using the probability proportional to size method, stratified by topology, sex, and registered year. The overall coding completeness of the registered and reabstracted data was 89.9% and 93.6%, respectively. In addition, the overall agreement rate among variables ranged from 84.7% to 99.6%, and  $\kappa$  coefficient ranged from 0.619 to 0.995. The misclassification among unilateral organs caused lower coding completeness and agreement rate of laterality coding. The completeness of current residency could be improved using the reabstracting method. The lowest agreement rate was found among various categories of diagnosis basis. Sex misclassification for male breast cancer was identified.

**Conclusion** The coding completeness and data reliability of the NCI hospital-based cancer registry met the standard in most critical variables. However, some challenges remain to improve the data quality. The reabstracting method could identify the critical points affecting the quality of cancer registry data.

© 2018 by American Society of Clinical Oncology Licensed under the Creative Commons Attribution 4.0 License

Anupong  
Sirirungreung

Rangsiya Buasom

Chuleeporn  
Jiraphongsa

Suleeporn Sangrajrang

Author affiliations and support information (if applicable) appear at the end of this article.

**Corresponding author:**  
Anupong Sirirungreung,  
MD, MPH, Phramongkut-  
kiao College of Medicine,  
315 Ratchawithi Rd,  
Ratchathewi District,  
Bangkok 10400, Thai-  
land; e-mail: anu.siri@  
pcm.ac.th.

## INTRODUCTION

According to the world cancer statistics report 2013 by the International Agency for Research on Cancer, global cancer burden increased between 2008 (12.7 million new cases and 7.6 million deaths) and 2012 (14.1 million new cases and 8.2 million cancer deaths).<sup>1</sup> In Thailand, cancer was the leading cause of death over the last decade. In 2014, 70,075 Thais died as a result of cancer.<sup>2</sup> The cancer registry is an essential part of cancer prevention and control. The cancer registry is an organization for the systematic collection, storage, analysis, interpretation, and reporting of data on patients with cancer.<sup>3</sup>

The data quality of the cancer registry is important to describe the extent of the cancer burden, to be a source of material for etiology studies, and to monitor and assess cancer prevention and control activities. Comparability, completeness, validity or accuracy, and timeliness are four dimensions to evaluate the data quality of the cancer registry.<sup>4,5</sup>

The two types of cancer registries defined by population are hospital-based and population-based. The hospital-based cancer registry gathers information on patients with cancer visited in a particular hospital, whereas the population-based cancer registry attempts to collect data

on patients with newly diagnosed cancer in a well-defined population.<sup>3</sup> Established in 1996, to date, the Thailand cancer registry has covered 389 hospitals from all regions and become a population-based cancer registry in some geographic areas.<sup>6</sup>

Few studies have been conducted regarding data quality of the cancer registry in Thailand. Between 1998 and 2000, Sriplung<sup>7</sup> found that only three of nine cancer registry centers met the standard for cancer incidence in five continents according to two indices: the percent of death certificates only and the percent of morphologically verified cases. In a capture-recapture study by Suwanrungruang et al,<sup>8</sup> the completeness of data across nine Thai cancer registries from 2003 to 2007 varied from 70% to 99.7%. In addition, Suwanrungruang et al<sup>8</sup> suggested that the mortality incidence ratio should not be used to evaluate the completeness of case ascertainment, because it does not reflect the quality of cancer registry procedures.<sup>9</sup> To address the cancer registry process in terms of data agreement with source of medical records, the reabstracting method could be deployed.<sup>4</sup>

The National Cancer Institute (NCI) in Bangkok, Thailand comprises functions of both hospital- and population-based cancer registries. The hospital-based cancer registry of NCI registers new cases of patients with cancer who visit the NCI. The patients of NCI come from many parts of Thailand. However, the population-based function of NCI's cancer registry collects data from other hospital within the Bangkok catchment area. In total, 2,916 to 3,917 new cancer cases annually were entered in the NCI hospital-based cancer registry.<sup>10</sup> Data entry is performed by well-trained cancer registry officers. The double-entry process for routine data entry is not used. However, a logical check using International Agency for Research on Cancer CanReg program coding<sup>11</sup> is manually run by programmers when combining data for annual reports. Moreover, a web application program, which automatically checks, was widely implemented after 2014. Thus, we conducted this cancer data quality study using the reabstracting method to evaluate coding completeness and data reliability of the NCI hospital-based cancer registry.

## METHODS

We retrieved the registered data and medical records concerning patients with colon, rectum, liver, lung, breast, cervix, and prostate cancer registered by the NCI hospital-based cancer registry between 2012 and 2014. Those topographies constituted the most common cancers among males and females in Thailand. The sample size was calculated based on a proportion to size calculation<sup>12</sup> using the proportion of non-Bangkok residents ( $P = 0.231$ ),<sup>10</sup> which contributed the largest number of samples to accomplish the study objectives. We added 10% to the sample size to compensate for any lost medical records, so the size totaled 971. Samples of records were then selected using the probability proportional to size method stratified by topology, sex, and registered year. We audited all cases of male breast cancer, because the number was small ( $n = 8$ ). Registered and reabstracted data were compared. The registered data comprised a set of patient data retrieved from the NCI hospital-based cancer registry database, which assisted the new linkage of identification numbers and marked patient identity variables. The reabstracted data comprised a set of data collected from the medical records by a new team of trained NCI audit staff. Three trained audit staff members from the national data audit units of NCI were deployed to reabstract the data according to national guidelines for cancer registry records. Reviewed and paper-recorded data were entered using Centers for Disease Control and Prevention Epi-Info, version 3.5.4. Double data entry was followed to recheck any data entry error.

The 14 variables included date of birth, sex, religion, citizenship registration province, current resident province, date of diagnosis, diagnosis basis, two-digit topology code, malignant behavior, laterality, grade, five-digit morphology code, life status, and last contact date, to be retrieved from registered data and reviewed. The missing data were defined as blank, unknown, ruled out cancer (for behavior variable), non-stage or not applicable (for grade variable), unavailable code 80003 (for morphology), unavailable code 809 (for topology), or unidentified data in medical records. To check the missing data, coding completeness was calculated as follows: coding completeness (%) =  $100 - (\text{number of missing value} / \text{total number of data elements})$ .

**Table 1.** Sampling Distribution of Reabstracted Data by Year of Diagnosis, Cancer Site, and Sex: The National Cancer Institute Hospital-Based Cancer Registry, Thailand, 2012 to 2014

Study Variable	Total NCI Database	Reabstracted Data
Year of diagnosis		
2012	1,751 (32.1)	313 (32.7)
2013	1,778 (32.6)	312 (32.6)
2014	1,933 (35.6)	332 (34.7)
Cancer site		
Colon	628 (11.5)	109 (11.4)
Rectum	377 (6.9)	66 (6.9)
Liver	659 (12.1)	114 (11.9)
Bronchus and lungs	849 (15.5)	147 (15.4)
Breast	2,145 (39.3)	382 (39.9)
Cervix	673 (12.3)	117 (12.2)
Prostate	131 (2.4)	22 (2.3)
Sex		
Male	1,720 (31.5)	306 (32.0)
Female	3,742 (68.5)	651 (68.0)
Total	5,462 (100)	957 (100)

NOTE. Data presented as No. (%).

Abbreviation: NCI, National Cancer Institute.

The data reliability was presented in two indices: the data agreement rate and Cohen's  $\kappa$  coefficient. The R program with package irr (Various Coefficients of Interrater Reliability and Agreement), version 0.84 was used to calculate the data agreement rate and unweighted  $\kappa$  coefficient of each variable except date. The missing data of each record were excluded from the analysis. The discrepancies of date variables were categorized as > 30 days, > 180 days, and > 365 days. Percent of nondiscrepancies of each category were calculated as: nondiscrepancies (%) = (number of reabstracted records with no discrepancies on date variable/total number of reabstracted records)  $\times$  100. Because NCI serves as an excellence center for cancer care in Thailand, some patients were referred from other hospitals. The diagnosis date was defined as that recorded in the referral paper. When the date was not recorded, or for unclear diagnosis, the following dates were used: biopsy or pathologic confirmation at NCI, cytology or imaging report, and physical examination by doctors. The life status and last contact date were excluded from calculation because of uncertain follow-up protocol. The stratified analysis of agreement rate and percent of nondiscrepancies was conducted

to identify the variability across the year of diagnosis and cancer site.

The data retrieval and reabstraction processes were conducted between August and September 2015. This study received approval from the Ethics Committee of NCI, Thailand (protocol No. 102\_2015RB\_IN432).

## RESULTS

From 2012 to 2014, 5,462 records of colon, rectum, liver, lung, breast, cervix, and prostate cancer were added in the NCI hospital-based cancer registry. A total of 957 records were reabstracted and compared with the registered records. Of 979 retrieved records, 22 were missing during the reabstracted process. The sampling distribution of NCI hospital-based cancer registry records, reabstracted records by year of diagnosis, cancer site, and sex are presented in [Table 1](#).

In total, 13,706 data elements (979 records  $\times$  14 data elements) from registered data and 13,398 data elements (957 records  $\times$  14 data elements) from reabstracted data were assessed for missing data. The overall coding completeness rates were 89.9% and 93.6%, respectively. A lower level of coding completeness was observed for grade, laterality, and morphology variables. As a result, the reabstracting process could improve the current resident province of the patients ([Table 2](#)). When we assumed that the missing current resident province was the same as citizenship registration province in the registered data, then the reabstracting process could improve 17 of 504 data elements (3.4%) of missing registered data. In all, 179 of 263 (68.1%) data sets on laterality among unilateral organs (ie, cervix, liver, rectum, colon, and prostate) were missing.

The overall agreement rate among variables ranged from 84.7% to 99.6%, and  $\kappa$  coefficient ranged from 0.619 to 0.995 ([Table 3](#)). The lowest overall agreement rate was observed for diagnosis basis (84.7%), followed by morphology code (88.4%), grade (89.3%), and laterality (89.8%).

The percent of nondiscrepancies regarding the diagnosis date following > 30-day criteria was 73.6%, which dramatically increased to 92.0% using > 180-day criteria. Consequently, the percent of nondiscrepancies concerning date

**Table 2.** Coding Completeness of Registered and Reabstracted Data: The National Cancer Institute Hospital-Based Cancer Registry, Thailand, 2012 to 2014

Variable	Registered Data (n = 979)		Reabstracted Data (n = 957)	
	Missing* (No.)	CC (%)	Missing* (No.)	CC (%)
Sex	0	100.0	0	100.0
Date of diagnosis	0	100.0	0	100.0
Behavior	5	99.5	2	99.8
Topology code	0	100.0	3	99.7
Current resident province	504	48.5	6	99.4
Citizenship registration province	2	99.8	8	99.2
Diagnosis basis	0	100.0	8	99.2
Date of birth	5	99.5	11	98.9
Religion	11	98.9	22	97.7
Life status	9	99.1	38	96.0
Last contact date	14	98.6	58	93.9
Laterality	263	73.1	115	88.0
Morphology code	167	82.9	183	80.9
Grade	404	58.7	404	57.8
Total	1,384	89.9	858	93.6

Abbreviation: CC, coding completeness.

\*Number of missing defined as: blank, unknown, ruled out cancer (for behavior variable), non-stage or not applicable (for grade variable), unavailable code 80003 (for morphology), unavailable code 809 (for topology), or unidentified data in medical record.

of birth were 86.0% and 87.7%, respectively (Table 4). Using the > 30-day criteria, we found that the lowest percent of nondiscrepancies were in regard to date of diagnosis in prostate (40.9%), breast (67.8%), colon (71.6%), and rectal (74.2%) cancers.

The agreement rate and percent of nondiscrepancies by year of diagnosis (2012 to 2014) across key study variables were similar (Fig 1). However, after stratifying by cancer site (Fig 2),

we found that the lowest overall agreement rates of diagnosis basis were observed for lung (59.3%), liver (70.2%), and prostate (77.2%) cancers. For the morphology code, the lowest overall agreement rates of diagnosis basis were for cervix cancer (71.4%), followed by liver (78.6%) and colon (87.9%) cancers. For grade, the lowest overall agreement rates of diagnosis basis were rectum cancer (83.0%), followed by breast (88.4%) and colon (89.2%) cancers. For laterality, the lowest overall agreement rates of diagnosis basis were in prostate cancer (42.9%), followed by liver (60.3%) and colon (79.2%) cancers.

The misclassification of sex among males with breast cancer in the registered data was identified. We found that five of eight were female by reabstracting results. The double check was conducted by reevaluating with medical records, and the results were the same.

## DISCUSSION

The quality of cancer registry data could affect interpretation of cancer situation and risk factor identification.<sup>13-16</sup> The reabstracting method was used to evaluate data agreement with source of medical records.<sup>4</sup> However, some variations, in

**Table 3.** Data Agreement Rate and  $\kappa$  Coefficient Between Registered and Reabstracted Data: The National Cancer Institute Hospital-Based Cancer Registry, Thailand, 2012 to 2014

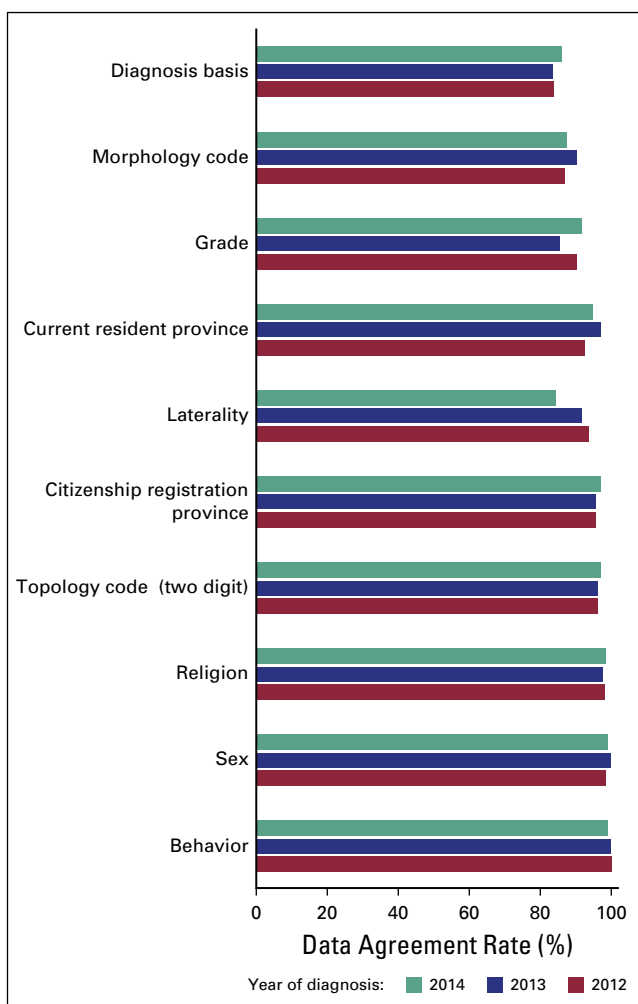
Variable	Observed (No.)	Agreement Rate (%)	$\kappa$ Coefficient
Diagnosis basis	949	84.7	0.619
Morphology code	765	88.4	0.839
Grade	486	89.3	0.830
Laterality	653	89.8	0.847
Current resident province	462	93.7	0.924
Citizenship registration province	948	96.2	0.957
Topology code (two digit)	953	96.5	0.955
Religion	924	98.3	0.961
Sex	957	99.2	0.981
Behavior*	950	99.6	—

\*The  $\kappa$  coefficient of behavior variable could not be computed because there was only one category identified in the registered data.

**Table 4.** Percent of Nondiscrepancies of Date Variables: The National Cancer Institute Hospital-Based Cancer Registry, Thailand, 2012 to 2014

Date Variable	Observed (No.)	Nondiscrepancy (%)
Date of diagnosis		
> 30 days	957	73.6
> 180 days	957	92.0
> 365 days	957	95.0
Date of birth		
> 30 days	941	86.0
> 180 days	941	87.7
> 365 days	941	97.1

**Fig 1.** Data agreement rate between registered and reabstracted data by year of diagnosis from the National Cancer Institute hospital-based cancer registry, Thailand, 2012 to 2014.



assessing the data agreement, were based on the data source compared. One group of studies used the existing database, such as cohort study, mortality data, and population census.<sup>13,15,17,18</sup> Another used medical records as the data source and deployed a well-trained audit to reabstract the data.<sup>19-21</sup> In this study, we

compared registry data with reabstracted data from medical records in a hospital.

Our study found that the overall coding completeness rate was good in most variables; however, some need to be improved. A 90% of coding completeness rate was recommended for survival analysis in one US study.<sup>13</sup> Age (date of birth) and sex variables of the NCI registered data met the US Cancer Statistics publication standard criteria.<sup>22</sup> However, four of 14 variables (laterality, current resident province, grade, and morphology code) displayed < 90% coding completeness. Some challenges of coding completeness need to be addressed. First, the laterality misclassification mostly resulted from unilateral organs coded as unknown. It could be improved using logical checking and recode. Second, the current resident province could be improved by medical record review only. However, a further exploration of similarity between citizenship registration and current resident was required. The coding completeness rates for some variables of reabstracted data were lower than the registered data. This could be attributed to handwriting interpretation difficulty and fading of printing over time.

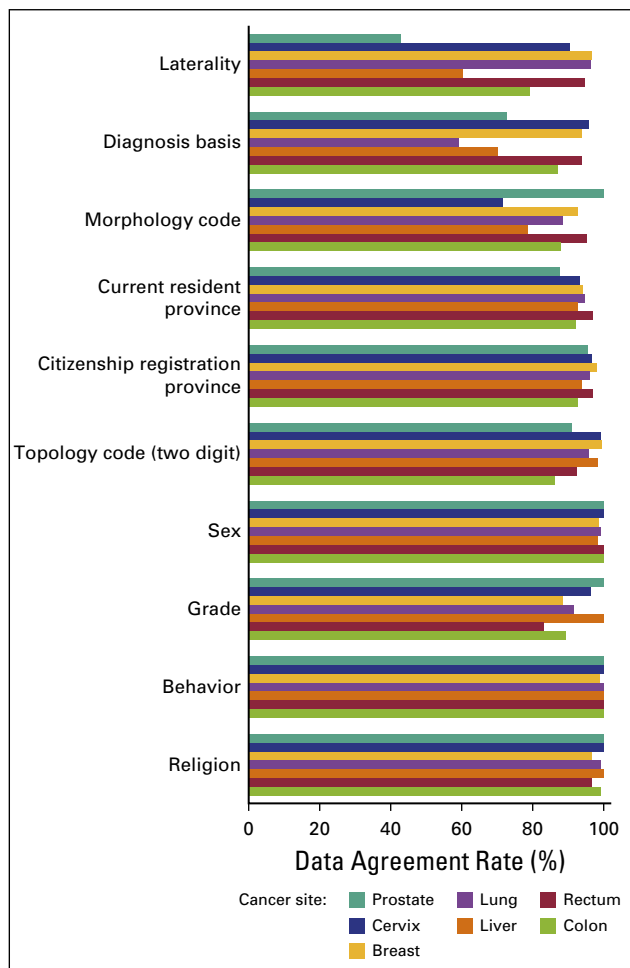
The high discrepancies of diagnosis date (> 30 days) might have occurred especially for the slowly progressing cancers and those requiring more diagnostic intervention. This was due to time lags between the first visit diagnosis date by clinical criteria and second visit confirmation by pathology results. However, this may also have resulted from not strictly relying on diagnosis-date criteria, because the staff may not have waited for the confirmed results. Waiting for complete data of diagnosis basis and diagnosis date might improve the quality of data. Furthermore, limitations of data retrieval may have occurred. Some medical records, especially hard copies of referral papers, may have been lost when returned to the local hospital.

Similar to other studies, in this study overall data reliability was good (ie, data agreement rate > 80% and κ coefficient of 0.6).<sup>20,21</sup> However, the data agreement rate in some countries could reach from 76.2% to 100%.<sup>19-21,23,24</sup> Opportunities to further improve data accuracy were identified. First, the NCI cancer registry record had various categories of diagnosis basis (ie, history, imaging, biochemistry test, cytology test, metastasis biopsy, and primary biopsy), which confuse



coders. A clearer definition and related evidence of each diagnosis basis category need to be considered. Some complex histologic tumors might be difficult to identify, such as those associated with ovarian cancer.<sup>17</sup> However, a subcategorized diagnosis basis may cause variation and discrepancies across tumor type. To answer the discrepancies for clinical diagnosis versus morphology, or morphology versus cytogenetics, requires more specialists' decisions (eg, a pathologist or oncologist), which lies beyond the capability of our cancer registry officers. In addition, some discrepancies were found between malignancy and primary tumor, which causes challenges in cancer diagnosis over time.<sup>25</sup> Second, miscoding of laterality in unilateral organs may have resulted from a variety of diagnosis records, such as the right lobe of the liver, right side of the colon, or left side of the prostate. However, the coders need to follow the standard of laterality identification provided by the

**Fig 2.** Data agreement rate between registered and reabstracted data by cancer site from the National Cancer Institute hospital-based cancer registry, Thailand, 2012 to 2014.



SEER program,<sup>26</sup> which objectively describes the paired organs.

Breast cancer in males is rare, and therefore misclassification by sex may distort the incident rate interpretation. A process of sex verification should be implemented when a male patient with breast cancer is identified.

Our study was subject to limitations. First, the internal validity across reabstracting team members was not assessed; however, we tried to limit this variation by training and providing a coding manual. Second, we found 22 medical records that could not be retrieved from the NCI Medical Record Department. Medical records of dead patients might have been pulled out from the documentation system and then brought together, making them difficult to sort and follow. In addition, some were lost in the referral process. However, these constituted a small number and did not disturb sampling distribution and result interpretation. Finally, other cancer sites excluded in this study may have more variability in agreement rate, such as nonsolid tumors, which use more complex morphology coding. Thus, the overall interpretation of this study should cautiously infer overall data reliability and coding completeness of the NCI hospital-based cancer registry.

The interpretation of agreement rate and  $\kappa$  coefficient should be a concern for variables that have large amounts of missing data. Although the missing data cannot directly be identified as discordant between original and reabstracted data, Krippendorff's  $\alpha$  might be recommended for this issue.<sup>27-30</sup> However, agreement rate and  $\kappa$  coefficient are mostly presented in the literature regarding cancer registry data quality.<sup>19-21,23,24</sup>

The process of data quality evaluation using various methods and indicators could emphasize an opportunity for improvement of the cancer registry. Furthermore, training registrars<sup>31</sup> and use of computer algorithm technology<sup>32,33</sup> are proven processes of data quality improvement. Therefore, the NCI has developed the Thai Cancer-Based Online web application to improve data quality and information sharing between cancer registry centers across the nation.<sup>34</sup>

In conclusion, between 2012 and 2014, the overall data quality of the Thai NCI in terms of coding completeness and reliability was good compared with international standards. The

reabstracting method provided more insights on data recoding protocol and adherence among coders. Additional improvement is recommended regarding the completeness of current resident status, defining of diagnosis date, diagnosis basis categorization, laterality coding of unilateral organs, and verification of rare male breast cancer. A double-check approach for data accuracy

should be considered to update more available data. An algorithm in computer programming technology needs to be deployed to pose reminders of rare cancers among a particular sex or age group.

DOI: <https://doi.org/10.1200/JGO.17.00147>  
Published online on [jgo.org](http://jgo.org) on April 12, 2018.

#### AUTHOR CONTRIBUTIONS

**Conception and design:** All authors

**Administrative support:** Anupong Sirirungreung, Rangsiya Buasom, Suleeporn Sangrajrang

**Provision of study material or patients:** Anupong Sirirungreung, Rangsiya Buasom

**Collection and assembly of data:** Anupong Sirirungreung, Rangsiya Buasom

**Data analysis and interpretation:** Anupong Sirirungreung, Rangsiya Buasom, Suleeporn Sangrajrang

**Manuscript writing:** All authors

**Final approval of manuscript:** All authors

**Accountable for all aspects of the work:** All authors

#### Rangsiya Buasom

No relationship to disclose

#### Chuleeporn Jiraphongsa

No relationship to disclose

#### Suleeporn Sangrajrang

No relationship to disclose

#### ACKNOWLEDGMENT

Supported by a grant from the collaboration program between the Ministry of Public Health of Thailand and the US Centers for Disease Control and Prevention. The contents are solely the responsibility of the authors and do not necessarily represent the views of The Task Force for Global Health, Training Programs in Epidemiology and Public Health Interventions Network, or the Centers for Disease Control and Prevention. We thank the National Cancer Institute auditing staff for their support with data retrieval and medical record reabstraction. We also thank the staff of the National Cancer Institute Medical Record Department for assisting in searching and following medical records. Finally, we thank Virginia Senkomago, MPH, PhD, Siddharudha Shivalli, MBBS, MD, Mona Saraiya, MD, MPH, and the staff of the Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention, and Health Promotion Centers for Disease Control and Prevention for providing technical consultation on cancer registry.

#### AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to [www.asco.org/rwc](http://www.asco.org/rwc) or [ascopubs.org/jco/site/ifc](http://ascopubs.org/jco/site/ifc).

#### Anupong Sirirungreung

No relationship to disclose

#### Affiliations

**Anupong Sirirungreung**, Phramongkutklao College of Medicine; **Rangsiya Buasom** and **Suleeporn Sangrajrang**, National Cancer Institute; and **Chuleeporn Jiraphongsa**, Field Epidemiology Training Program, Bangkok, Thailand.

#### REFERENCES

1. International Agency for Research on Cancer: Latest world cancer statistics Global cancer burden rises to 14.1 million new cases in 2012: Marked increase in breast cancers must be addressed. [http://www.iarc.fr/en/media-centre/pr/2013/pdfs/pr223\\_E.pdf](http://www.iarc.fr/en/media-centre/pr/2013/pdfs/pr223_E.pdf)
2. Bureau of Policy and Strategy Ministry of Public Health: Statistical Thailand 2015 [in Thai]. [http://bps.moph.go.th/new\\_bps/sites/default/files/statistic58.pdf](http://bps.moph.go.th/new_bps/sites/default/files/statistic58.pdf)
3. International Agency for Research on Cancer: Cancer epidemiology: Principles and methods. <http://www.iarc.fr/en/publications/pdfs-online/epi/cancerepi/>
4. Bray F, Parkin DM: Evaluation of data quality in the cancer registry: Principles and methods. Part I: comparability, validity and timeliness. *Eur J Cancer* 45:747-755, 2009
5. Parkin DM, Bray F: Evaluation of data quality in the cancer registry: Principles and methods Part II. Completeness. *Eur J Cancer* 45:756-764, 2009

6. National Cancer Institute: Cancer in Thailand, 2010-2012, Volume VIII. Imsamran W, Chaiwerawattana A, Wiangnon S, et al (eds).[http://www.nci.go.th/th/File\\_download/NciCancerRegistry/CancerinThailand8.pdf](http://www.nci.go.th/th/File_download/NciCancerRegistry/CancerinThailand8.pdf)
7. Sriplung H: Thailand cancer registry data quality and cancer situation with associated factors from multiple data sources [in Thai].[http://www.hiso.or.th/hiso/picture/reportHealth/n\\_report1.pdf](http://www.hiso.or.th/hiso/picture/reportHealth/n_report1.pdf)
8. Suwanrungruang K, Sriplung H, Attasara P, et al: Quality of case ascertainment in cancer registries: A proposal for a virtual three-source capture-recapture technique. *Asian Pac J Cancer Prev* 12:173-178, 2011
9. Suwanrungruang K, Sriplung H, Temiyasathit S, et al: Appropriateness of the standard mortality/incidence ratio in evaluation of completeness of population-based cancer registry data. *Asian Pac J Cancer Prev* 12:3283-3288, 2011
10. National Cancer Institute: Hospital-based cancer registry annual report 2012. [http://www.nci.go.th/th/File\\_download/NciCancerRegistry/Hospital-BasedNCI2012Total.pdf](http://www.nci.go.th/th/File_download/NciCancerRegistry/Hospital-BasedNCI2012Total.pdf)
11. International Association of Cancer Registries: Software for cancer registries: CanReg5. [http://www.iacr.com.fr/index.php?option=com\\_content&#x0026;view=article&#x0026;id=9:canreg5&#x0026;catid=68&#x0026;Itemid=445](http://www.iacr.com.fr/index.php?option=com_content&#x0026;view=article&#x0026;id=9:canreg5&#x0026;catid=68&#x0026;Itemid=445)
12. Sullivan KM, Pezzullo JC, Dean AG, et al: Sample size for a proportion or descriptive study: Open source statistics for public health. <http://www.openepi.com/SampleSize/SSPropor.htm>
13. Wilson RJ, O'Neil ME, Ntekop E, et al: Coding completeness and quality of relative survival-related variables in the National Program of Cancer Registries Cancer Surveillance System, 1995-2008. *J Registry Manag* 41:65-71, 2014
14. Møller H, Richards S, Hanchett N, et al: Completeness of case ascertainment and survival time error in English cancer registries: Impact on 1-year survival estimates. *Br J Cancer* 105:170-176, 2011
15. Peres SV, Latorre MD, Tanaka LF, et al: Quality and completeness improvement of the Population-based Cancer Registry of São Paulo: Linkage technique use. *Rev Bras Epidemiol* 19:753-765, 2016
16. Brenner H, Hakulinen T: Population-based monitoring of cancer patient survival in situations with imperfect completeness of cancer registration. *Br J Cancer* 92:576-579, 2005
17. Mukherjee AK, Leck I, Langley FA, et al: The completeness and accuracy of health authority and cancer registry records according to a study of ovarian neoplasms. *Public Health* 105:69-78, 1991
18. Seneviratne S, Campbell I, Scott N, et al: Accuracy and completeness of the New Zealand Cancer Registry for staging of invasive breast cancer. *Cancer Epidemiol* 38:638-644, 2014
19. Shimakawa Y, Bah E, Wild CP, et al: Evaluation of data quality at the Gambia national cancer registry. *Int J Cancer* 132:658-665, 2013
20. Elbasmi AA, Fayaz MS, Al-Mohanadi S, et al: Reliability of the Kuwait Cancer Registry: A comparison between breast cancer data collected by clinical oncologists and registry staff. *Asian Pac J Cancer Prev* 11:735-738, 2010
21. Gulliford MC, Bell J, Bourne HM, et al: The reliability of cancer registry records. *Br J Cancer* 67:819-821, 1993
22. Centers for Disease Control and Prevention: United States Cancer Statistics publication criteria. [http://www.cdc.gov/cancer/npcr/uscs/technical\\_notes/criteria.htm](http://www.cdc.gov/cancer/npcr/uscs/technical_notes/criteria.htm)
23. Larsen IK, Småstuen M, Johannesen TB, et al: Data quality at the Cancer Registry of Norway: An overview of comparability, completeness, validity and timeliness. *Eur J Cancer* 45:1218-1231, 2009
24. Thoburn KK, German RR, Lewis M, et al: Case completeness and data accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries. *Cancer* 109:1607-1616, 2007



25. Vogt A, Schmid S, Heinimann K, et al: Multiple primary tumours: Challenges and approaches, a review. *ESMO Open* 2:e000172, 2017
26. National Cancer Institute: 2012 SEER Program Coding and Staging Manual: Description of this neoplasm: Laterality. NIH Publication number 12-5581, Bethesda, MD. 2012., 67-69 pp.[http://seer.cancer.gov/archive/manuals/2012/SPCSM\\_2012\\_maindoc.pdf](http://seer.cancer.gov/archive/manuals/2012/SPCSM_2012_maindoc.pdf)
27. Hallgren KA: Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor Quant Methods Psychol* 8:23-34, 2012
28. Zapf A, Castell S, Morawietz L, et al: Measuring inter-rater reliability for nominal data: Which coefficients and confidence intervals are appropriate? *BMC Med Res Methodol* 16:93, 2016
29. Viera AJ, Garrett JM: Understanding interobserver agreement: The kappa statistic. *Fam Med* 37:360-363, 2005
30. McHugh ML: Interrater reliability: The kappa statistic. *Biochem Med (Zagreb)* 22:276-282, 2012
31. Corley B, Ramadan M, Smith BL, et al: Measuring the effect of improved medical facilities and focused training on data quality and completeness: An example from the Gharbiah population-based cancer registry, Egypt. *J Registry Manag* 42:86-90, 2015
32. Contiero P, Tittarelli A, Maghini A, et al: Comparison with manual registration reveals satisfactory completeness and efficiency of a computerized cancer registration system. *J Biomed Inform* 41:24-32, 2008
33. al-Haddad BJS, Jedy-Agba E, Oga E, et al: Comparability, diagnostic validity and completeness of Nigerian cancer registries. *Cancer Epidemiol* 39:456-464, 2015
34. National Cancer Institute, Thailand: Thai Cancer-Based Report. <http://tcb.nci.go.th/CWEB/cwebBase.do?mode=initialApplication>