Journal of the Formosan Medical Association 120 (2021) 2037—2041

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jfma-online.com

Short Communication

Accuracy of long-form data in the Taiwan cancer registry

Chia-Wen Kao [a,1], Chun-Ju Chiang [b,c,1], Li-Ju Lin [d], Chiao-Wen Huang [d], Wen-Chung Lee [a,b,c,e,*], Ming-Yang Lee [f,], the Taiwan Society of Cancer Registry Expert Group[2]**

[a] *Department of Public Health, College of Public Health, National Taiwan University, Taipei, Taiwan*
[b] *Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan*
[c] *Taiwan Cancer Registry, Taipei, Taiwan*
[d] *Health Promotion Administration, Ministry of Health and Welfare, Taipei, Taiwan*
[e] *Innovation and Policy Center for Population Health and Sustainable Environment, College of Public Health, National Taiwan University, Taipei, Taiwan*
[f] *Department of Internal Medicine, Division of Hemato-oncology, Chia-Yi Christian Hospital, Chiayi, Taiwan*

Received 21 January 2021; received in revised form 19 April 2021; accepted 26 April 2021

KEYWORDS
Cancer registry;
Quality indicator;
Accuracy rate

The Taiwan Cancer Registry (TCR) is a nationwide population-based registry that collects the data of patients with newly diagnosed cancer from hospitals with ≥50 beds. TCR data are high quality in terms of completeness and timeliness. However, accuracy is also a crucial quality indicator. This study evaluated the accuracy rates of selected 55 major items in the long-form TCR data between 2014 and 2016 with 700 reported cases randomly selected from 25 long-form-reporting hospitals. We calculated the accuracy rates of the reported data by employing a reabstracted chart review. Among the 55 items, the accuracy rates of 38 (69%) were at least 95%, those of 10 (18%) were between 90% and 95%, those of 5 (9%) were between 85% and 90%, and the remaining 2 (4%) were between 80% and 85%. This demonstrates a high degree of accuracy in the TCR long-form data.

* Corresponding author. Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Room 536, No. 17 Xuzhou Rd., Taipei 100, Taiwan.
** Corresponding author. Department of Internal Medicine, Division of Hemato-oncology, Chia-Yi Christian Hospital, No. 539 Zhong-Xiao Road, Chiayi, Taiwan.
E-mail addresses: wenchung@ntu.edu.tw (W.-C. Lee), 05825@cych.org.tw (M. Lee).
[1] **Equal contribution first authors**: Chia-Wen Kao and Chun-Ju Chiang contributed equally to this work.
[2] List of members of the Taiwan Society of Cancer Registry Expert Group are listed at Appendix B.

https://doi.org/10.1016/j.jfma.2021.04.022

The Taiwan Cancer Registry (TCR) is a nationwide population-based registry that provides critical data on cancer incidence, care, and survival in Taiwan. The TCR has collected the data of patients with newly diagnosed cancer from hospitals with ≥50 beds in Taiwan since 1979.[1,2] Each year, qualified cancer registrars at the reporting hospitals identify new cancer cases and perform online logic checks for potential errors before they submit the data to the TCR. The TCR central office uses standardized algorithms to validate the received data. Any questionable data are returned to the reporting hospitals for reconfirmation.[3] The completeness, quality, and timeliness of TCR data have been improved. In 2016, the completeness of TCR data increased to 98.4%, the proportion of cases with morphological verification of cancer diagnoses increased to 93.0%, the mortality—incidence ratio decreased to 45.1%, the percentage of cases with only a death certificate decreased to 0.9%, and the interval between the date of diagnosis and date of reporting decreased to 14 months. These values indicate the high quality of the TCR data.[1–3]

The TCR data are high quality in terms of completeness and timeliness.[1–3] However, accuracy is also a crucial quality indicator for registry data. The accuracy rates of specific items in cancer registry data have been evaluated. For example, in the Tennessee Cancer Registry, the accuracy rate of "surgery of the primary site" from the Commission on Cancer (CoC) facilities was lower than that from non-CoC facilities.[4] In Scotland, "grade/differentiation" in cancer registration data was less reliable than demographic and diagnostic data were.[5] In the TCR, cancer type and the experience of cancer registrars affect the accuracy rate of items related to the first course of cancer treatment, and the cancer caseloads of hospitals affect the accuracy rate of items related to cancer staging.[6,7]

This study comprehensively evaluated the accuracy rates of selected 55 major items in the long-form registry data of the TCR database.

We collected cases from the long-form TCR data between 2014 and 2016. During this period, a total of 88, 91, and 94 hospitals reported cases to the TCR by using the long-form format for 16 major cancers in 2014, 2015, and 2016, respectively. The nearly 100 remaining hospitals used the short-form format. By using purposive sampling (Fig. 1), we selected 25 long-form reporting hospitals in total for this study. Among the 25 hospitals, 22 were selected because they had high (>2%) error rates for logic checks in data reporting, and the remaining three were selected because they were new to the long-form reporting system during the study period. For each hospital, 28 reported cases were randomly selected for on-site visits and chart review. Finally, the data of 700 cases were

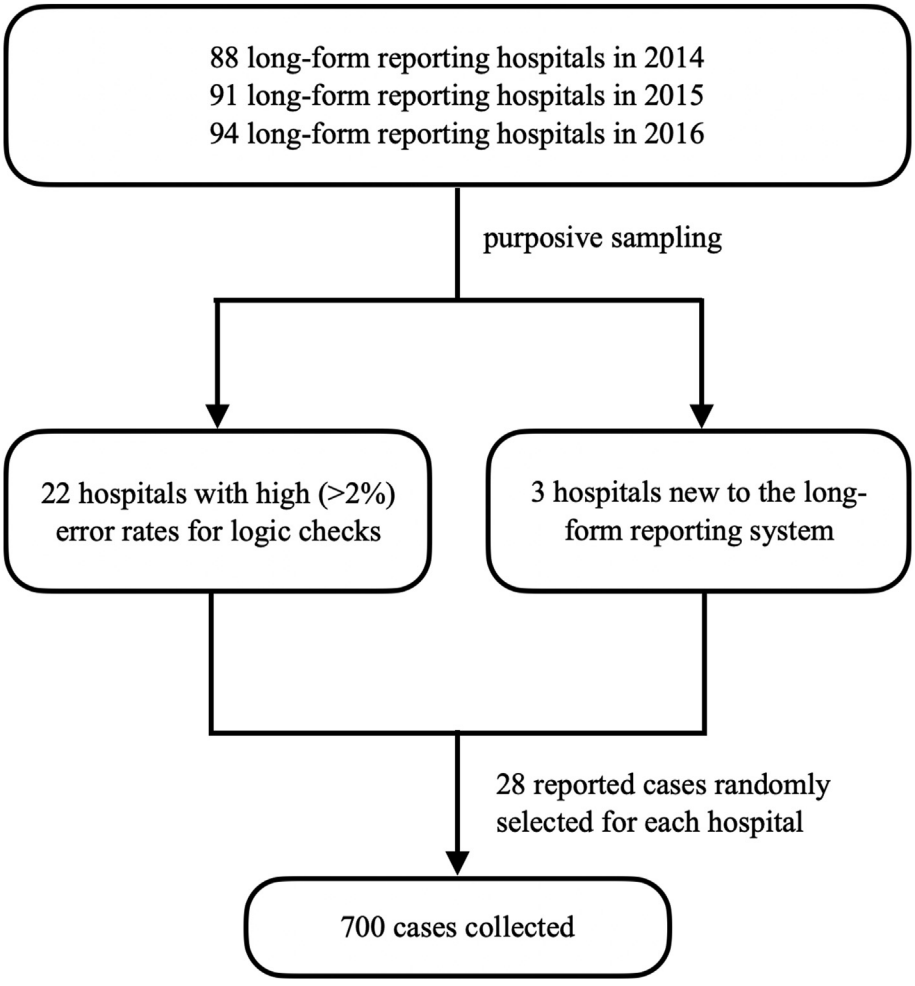

Figure 1 The two-level sampling process.

Table 1A Accuracy rates of items related to cancer identification and staging.

Item	Accuracy rate	95% confidence interval (right-sided)
Sequence number[a]	98.9%	>98.0%
Class of case[b]	95.1%	>93.6%
Date of first contact	86.7%	>84.4%
Date of initial diagnosis	82.7%	>80.2%
Primary site	92.3%	>90.4%
Histology	95.1%	>93.6%
Behavior code	99.7%	>99.1%
Grade/differentiation	95.1%	>93.6%
Diagnostic confirmation	97.0%	>95.7%
Date of first microscopic confirmation	90.4%	>88.4%
Tumor size	82.3%	>79.7%
Number of regional lymph nodes examined	98.0%	>96.9%
Number of regional lymph nodes positive	98.7%	>97.8%
Date of surgical diagnostic and staging procedure	87.0%	>84.7%
Surgical diagnostic and staging procedure at other facility	96.3%	>94.9%
Surgical diagnostic and staging procedure at this facility	90.1%	>88.1%
Clinical tumor (T)	87.4%	>85.2%
Clinical node (N)	92.3%	>90.4%
Clinical metastasis (M)	94.9%	>93.3%
Clinical stage group[c]	88.9%	>86.7%
Pathologic T	96.0%	>94.6%
Pathologic N	95.9%	>94.4%
Pathologic M	96.7%	>95.4%
Pathologic stage group[c]	94.1%	>92.5%
Clinical other stage group[d]	98.3%	>97.2%
Pathologic other stage group[d]	98.6%	>97.6%

[a] Order of malignant neoplasms the patient has had over their lifetime.
[b] Class of case reflects the hospital's role in managing cancer, and its grouping is based on the location of diagnosis and first course of treatment.
[c] Anatomic extent of disease based on the American Joint Committee on Cancer (AJCC) tumor, node, metastasis staging system.
[d] Non-AJCC cancer staging system for specific cancer sites.

collected. We then calculated the accuracy rates of the reported data of selected 55 long-form registry items in total, with the reabstracted data from the chart review serving as the gold standard. We also calculated the lower limit of the right-sided 95% confidence interval of the accuracy rate based on the binomial test (the exact method).

Because of the purposive sampling method (selecting hospitals with high error rates for logic checks in data reporting or hospitals that were new to the long-form reporting system), the accuracy rates calculated in this study were expected to be lower than the national average. Nevertheless, the lower limits of the right-sided 95% confidence intervals that we calculated in this study provide the conservative lower bounds of the accuracy rates.

Table 1(A) presents the accuracy rates of items related to cancer identification and staging. Among the 26 items, the accuracy rates of 14 (54%) were ≥95%, those of 6 (23%) were between 90% and 95%, and those of 6 (23%) were <90%. Items with a conservative lower bound of <85% included "date of surgical diagnostic and staging procedure" (84.7%), "date of first contact" (84.4%), "date of initial diagnosis" (80.2%), and "tumor size" (79.7%).

Table 1(B) presents the accuracy rates of items related to the first course of treatment. Among the 29 items, the accuracy rates of 24 (83%) were ≥95%, those of 4 (14%) were between 90% and 95%, and that of 1 (3%) was <90%. No items had a conservative lower bound of <85%. The conservative lower bounds of five radiotherapy-related items exceeded 99%.

Among all items in the TCR data, the accuracy rate of "tumor size" was the lowest (82.3%; conservative lower bound: 79.7%). Tumor size tended to be underestimated in the registry; the proportion of cases with underestimated tumor size (5.3%) was larger than of cases with overestimated tumor size (3.7%; Appendix Table 1). If a 10-mm error in tumor size is allowed, then the accuracy rate increases to 87.4% (conservative lower bound: 85.1%; Appendix Table 2). The accuracy rate of "date of initial diagnosis" was the second lowest (82.7%; conservative lower bound: 80.2%). The date of initial diagnosis tended to be recorded in the registry as later than it should have been; the proportion of cases with a late date of initial diagnosis (10.2%) was larger than that of cases with an early date of initial diagnosis (3.3%; Appendix Table 3). If a 3-day error in the date of initial diagnosis is allowed, then the accuracy rate increases to 87.3% (conservative lower bound: 85.0%; Appendix Table 4).

In this study, the accuracy rate of tumor size was the lowest among all items. The coding manual for the TCR specifies that if a patient has more than one tumor size record in the registry (such as records obtained through different imaging modalities), the registrar responsible should report the largest tumor size of all the records. However, registrars may often fail to retrieve all records regarding tumor size in practice because of limited time and the considerable effort involved. This helps explain why tumor size is often underestimated in the registry.

The date of initial diagnosis had the second-lowest accuracy rate and tended to be recorded as later than it should have been. According to the TCR coding manual, the date of initial diagnosis is that recorded by a physician for tumor diagnosis confirmed clinically or microscopically. A registrar may report the date of the pathological confirmation of cancer. However, if a physician states that upon retrospection, a patient had cancer at an earlier date, this earlier date should be reported as the date of initial diagnosis for the patient. Registrars may report the date when the test result the diagnosis was based on was obtained when in fact, it should be the date that the test was performed. This explains why the date of initial diagnosis was often late in the registry. The results indicate that if a 3-day error is allowed, the accuracy rate of the date of initial

Table 1B Accuracy rates of items related to first course of treatment.

Item	Accuracy rate	95% confidence interval (right-sided)
Date of first course of treatment	92.6%	>90.7%
Date of first surgical procedure	95.0%	>93.4%
Date of most definitive surgical resection of the primary site	96.0%	>94.6%
Date of chemotherapy started at this facility	94.3%	>92.6%
Date of hormone therapy started at this facility	97.9%	>96.7%
Date of immunotherapy started at this facility	99.4%	>98.7%
Date of targeted therapy started at this facility	97.7%	>96.5%
Surgical procedure of primary site at this facility	89.0%	>86.9%
Chemotherapy at this facility	92.4%	>90.6%
Hormone therapy at this facility	98.6%	>97.6%
Immunotherapy at this facility	99.4%	>98.7%
Targeted therapy at this facility	98.0%	>96.9%
Summary of radiotherapy target volume[e]	98.1%	>97.1%
Radiotherapy modality	98.9%	>98.0%
Date radiotherapy started	98.4%	>97.4%
Date radiotherapy ended	97.9%	>96.7%
Location of radiotherapy	98.3%	>97.2%
External beam radiotherapy technique	94.3%	>92.6%
Target sites for the highest radiation dose of the clinical target volume	96.9%	>95.5%
Highest radiation dose for the clinical target volume (cGy)	97.0%	>95.7%
Number of fractions for the highest radiation dose for the clinical target volume	96.9%	>95.5%
Target sites for the lower radiation dose for the clinical target volume	97.4%	>96.2%
Lower radiation dose for the clinical target volume (cGy)	97.0%	>95.7%
Number of fractions for the lower radiation dose for the clinical target volume	97.4%	>96.2%
Other radiotherapy modality	99.9%	>99.3%
Other radiotherapy technique	100.0%	>99.6%
Target of other radiotherapy	100.0%	>99.6%
Dose for the clinical target volume of other radiotherapy	99.9%	>99.3%
Number of fractions for the clinical target volume of other radiotherapy	99.9%	>99.3%

[e] Coverage of radiation target volume for radiotherapy in the first course of treatment.

diagnosis remains <90%. This may jeopardize the survival analysis of patients with cancer when the date of initial diagnosis is used as the starting point for follow-up and the focus is on short-term survival.

Funding

This paper was supported by grants from the Health Promotion Administration, Ministry of Health and Welfare (A1081115 Tobacco Health and Welfare Taxation) in Taiwan, the Ministry of Science and Technology in Taiwan (MOST 108-3017-F-002-001, MOST 108-2314-B-002-127-MY3), and the Innovation and Policy Center for Population Health and Sustainable Environment (Population Health Research Center, PHRC) from Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan (NTU-109L900308). No additional external funding was received for this study. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Ethical approval

The study was reviewed and approved by the Research Ethics Committee of the National Taiwan University (IRB No. 202011HM001).

Declaration of competing interest

The authors have no conflicts of interest relevant to this article.

Appendix B. List of members of the Taiwan Society of Cancer Registry Expert Group

Dr. Shiau Cheng-Yi (Division of Radiation Oncology, Department of Oncology, Taipei Veterans General Hospital, Taipei) and all senior cancer registrars, including Hsiu-Ling Lin, Mei-Man Lin, Yi-Ping Wang (National Taiwan University Hospital, Taipei), Ming-Ling Chen, Kuei-Chih Lin, Yueh-Ying Shen, Shu-Ching Wu (Taipei Veterans General Hospital, Taipei), Shu-Chen Cheng (Cathay General Hospital, Taipei), Pei-Lin Chen, Yen-Chun Lin (Koo Foundation Sun Yat-Sen Cancer Center, Taipei), Su-Lan Wang (Keelung Chang Gung Memorial Hospital, Keelung), Chin-

Ying Cheng (Taipei Tzu Chi Hospital, New Taipei), Chiao-Min Chen, Shu-Chen Chen, Hui-Wen Cheng, Liang-Yu Chiang , Wei-Ling Kao (LinKou Chang Gung Memorial Hospital, Taoyuan), Liao,Su-Chien (Chung Shan Medical University Hospital, Taichung), Chia-Fen Chang , Tsai-Chieh Chen, Wei-Ling Liau , Chia-Ling Lin (China Medical University Hospital, Taichung), Chia-Ling Lee, Yueh-Yun Suen, Chen-His Wang (Taichung Veterans General Hospital, Taichung), Win-Jieh Lai (Taichung Tzu Chi Hospital, Taichung), Tsai-Yun Huang, Yu-Ru Lin (Changhua Christian Hospital, Changhua), Hui-Yu Kao, Li-Hua Wei (St. Martin De Porres Hospital, Chiayi), Ching-Chin Huang (Chi Mei Medical Center, Tainan), Mei-Lin Kuo (National Cheng Kung University Hospital, Tainan), Chiao-Pin Huang (Kaohsiung Medical University Chung-Ho Memorial Hospital, Kaohsiung), Shiau-Wei Chen (E-Da Hospital, Kaohsiung), Li-Fang Chiu, Hui-Ting Chang (E-Da Hospital, Kaohsiung), Ling-Chu Lin (Hualien Tzu Chi Hospital, Hualien).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jfma.2021.04.022.

References

1. Chiang CJ, Wang YW, Lee WC. Taiwan's nationwide cancer registry system of 40 years: past, present, and future. *J Formos Med Assoc* 2019;**118**(5):856—8.
2. Taiwan Cancer Registry. *Taiwan cancer registry reporting manual*. 2020.
3. Chiang CJ, You SL, Chen CJ, Yang YW, Lo WC, Lai MS. Quality assessment and improvement of nationwide cancer registration system in Taiwan: a review. *Jpn J Clin Oncol* 2015;**45**(3):291—6.
4. Kiernan CM, Whiteside MA, Solorzano CC. Cancer registries: can we improve the quality of thyroid cancer data? *Ann Surg Oncol* 2017;**24**:1202e7.
5. Brewster DH, Stockton D, Harvey J, Mackay M. Reliability of cancer registration data in Scotland, 1997. *Eur J Cancer* 2002 Feb;**38**(3):414—7. https://doi.org/10.1016/s0959-8049(01)00385-9. PMID: 11818208.
6. Cheng CY, Chiang CJ, Hsieh CH, Chang YK, Lai MS. Is quality of registry treatment data related to registrar experience and workload? A study of Taiwan cancer registry data. *J Formos Med Assoc* 2018;**117**(12):1093—100.
7. Cheng CY, Chiang CJ, Hsiao JK, Lai MS. Are hospital cancer caseloads related to the validity of staging data reported? A lesson from National Cancer Registry in Taiwan. *Jpn J Clin Oncol* 2017 Jan;**47**(1):18—24.