

## Quality assessment of pathologic data in cancer registry centers based on ICD-O-3

Raziehsadat Mousavi (MSc)<sup>1</sup>  
 Ghahraman Mahmoudi (PhD)<sup>1</sup>  
 Hossein-Ali Nikbakht (PhD)<sup>2</sup>  
 Mohammad Ali Jahani (PhD)<sup>2\*</sup>

1. Hospital Administration  
 Research Center, Sari Branch,  
 Islamic Azad University, Sari, Iran  
 2. Social Determinants of Health  
 Research Center, Health Research  
 Institute, Babol University of  
 Medical Sciences, Babol, Iran

**\* Correspondence:**

**Mohammad Ali Jahani**, Social  
 Determinants of Health Research  
 Center, Health Research Institute,  
 Babol University of Medical  
 Sciences, Babol, Iran

**E-mail:** drmajahani@yahoo.com

**Tel:** 0098 1132190624

**Fax:** 0098 1132190624

### Abstract

**Background:** Prerequisite for achieving the goals of the registration program is the existence of valid and accurate data, and the usability of this data is possible if they are coded correctly. This study assesses the quality of pathological data of the population-based cancer registration centers based on ICD-O-3.

**Methods:** This applied study was performed descriptively and retrospectively. The study population included 20129 pathology reports sent to the population-based cancer registration center of Mazandaran Province during 2018-2020. A total of 2015 out of 2050 samples of the received reports were examined according to stratified random sampling method. A researcher checklist was made to collect the data, and STATA 13 and Cohen's Kappa agreement coefficient were used to analyze the data.

**Results:** Among the 2015 reports of pathology, 1114 (55.3%) pathology reports were related to government centers, (42.9%) 865 cases were registered with their topographic code, morphology and behavior. Based on the registration of the exact topographic code, the kappa coefficient and the total agreement were 0.266 and 27.70%, respectively. Kappa coefficient in all received reports and reports with topographic code was 0.346 and 0.906, respectively. In the reports with topographic code, the most reports of cancers were related to cancers of the gastrointestinal organs (97.6%) 246.

**Conclusion:** The accuracy of the codes given in the pathology centers in terms of topographic, morphological, behavioral and grade codes based on the percentage of agreement with the coding was above average, which were higher in governmental centers and also in some cancers.

**Keywords:** Neoplasms, Grade, Assessment, Accuracy, Reliability

### Citation:

Mosavi R, Mahmoudi G, Nikbakht HA, Jahani MA. Quality assessment of pathologic data in cancer registry centers based on ICD-O-3. Caspian J Intern Med 2022; 13(3): 589-598.

Cancer is one of the most common non-communicable diseases and one of the most common diseases that in recent decades has been shocking statistics of mortality (1, 2). In Iran, cancer is one of the most common non-communicable diseases and after heart disease and road accidents is considered as the third leading cause of death (1), which makes it increasingly important for prevention, timely and correct diagnosis and finally treatment of this disease. Prevention and reduction of cancer cases requires the existence of cancer control programs, that the cancer registration system is one of the essential components of cancer control programs (2). The use of databases for information retrieval, epidemiological studies and management decisions is subject to correct clinical coding. Therefore, the accuracy of the coding of medical records is essential. As a result, the quality of the coded information must be evaluated, and these are the steps taken to ensure the quality of the coding.

**Received:** 18 Jan 2022

**Revised:** 1 April 2022

**Accepted:** 27 April 2022



Existence of coding error leads to incorrect classification of diseases and consequently incorrect statistics. Relying on data whose classification quality is poor will pose risks to both care providers and managers in planning and policymaking, education and research, care delivery, and so on. By identifying the factors related to the accuracy of disease coding, the necessary information can be provided to perform useful and effective interventions to improve the quality of disease information (3, 4).

The Cancer Registry Program is an information management system that collects, stores, manages, analyzes, and reports data on cancer patients and deaths from the disease (5, 6). On the other hand, the use of this information is possible when this information is properly organized, classified and coded (7, 8).

In recent years, the process of coding medical records due to communication closeness to care quality issues and reported data has become increasingly important nationally and internationally (8, 9). The coding accuracy of medical records is essential, and the quality of the coded information must be evaluated to ensure quality (3, 10). To integrate the tumor coding process into cancer registration programs, the ICD-O International Coding System has been designed by the International Agency for Research on Cancer. In the National Cancer Registration Program of Iran, the ICD-O-3 coding system is used to classify and encode tumor information (11).

Most of the studies performed on the evaluation of coding quality and accuracy show that the accuracy of coding of diagnoses is at a low level. In a study conducted in Germany, the reliability of coding of diagnoses was evaluated at a low to moderate level (12).

The results of the study of Caski showed that the USA health care system used diagnostic codes for billing and reimbursement in its research and showed that all cancer records in these countries are high standard in terms of completeness and accuracy (13).

The results of the study of Feiz showed that the average of accuracy of diagnostic accuracy was 80.3% and the average of accuracy of treatment measures was 84.2%. The reported accuracy also indicates that the data collected are typically strong enough to be used for research and managerial decision-making (14).

Therefore, considering the importance of coding information on cancer patients and the need for accuracy and precision of these codes, and considering that so far few studies have been done to evaluate and check the accuracy of

information and codes of cancer patients' records and the status of information registration and information accuracy and the codes assigned to the pathology reports sent to the population-based cancer registration centers have not been reviewed, this study is done to determine the status of data registration of pathology reports sent to the population-based cancer registry center of Mazandaran province and qualitative evaluation of pathological data based on the international classification system of cancer diseases to identify probable errors and provide solutions to eliminate them to improve the quality of the registration data.

## Methods

This applied study was performed descriptively and retrospectively, after obtaining the code of ethics at IR.IAU.CHALUS.REC.1399.025 from Islamic Azad University and receiving a letter of introduction. The study population included, 20129 pathology reports sent by 32 government centers and 73 private centers active in 14 cities of the province, which were sent to the Cancer Registration Center during the years 2018-2020. Using Cochran's formula and also considering the confidence level of 0.95, study power of 0.90 and acceptable error of 0.03 for the ratio, 1900 samples were needed to check the error between the coders of the pathology centers and the coder of the cancer registration center.

Therefore, about 10% of the samples of each year (three years studied) were considered as the final sample using stratified random sampling method. In such a way that by using proportional allocation and in proportion to the weight of the type of pathology center (government or private) in the research community, the sample size was determined for each year. Finally, 2050 pathology reports were reviewed in this study. 12 reports of them were received from out-of-university pathology centers to cancer registration centers and 23 reports were benign cancers, and they were excluded from the study, thus 2015 reports were considered in the final analysis.

A researcher-made checklist was used to collect data, experts and thinkers were used to assess the validity of the checklist, and after corrections, they were approved. By assigning a record code to each report, all patient information was published without mentioning the names and identities of patients and was collected while maintaining the principles of security and confidentiality.

Next, by examining the site of the cancer as well as the histology and behavior of the cancer cell announced in the pathology report and selecting the site and the type of cancer, referring to the cancer coding book and according to the existing rules, the necessary codes for the organ with the tumor (topography) including cancer code and grouping, tumor histology type (morphology), tumor behavior (benign, in situ or malignant) and cancer tumor grade (grade) were extracted from the book and recorded in the relevant checklist. The accuracy of the codes was assessed using the Cohen's Kappa agreement coefficient. In general, this index shows the degree of agreement of the two evaluators (excluding chance) on a two-state trait.

Kappa coefficient and statistical analysis based on it are numerical values between -1 to +1, which the closer to +1 indicates the existence of a proportional and direct agreement. Values close to -1 indicate inverse agreement, and values close to 0 indicate disagreement. Kappa coefficient, measures reliability of agreement of different persons regarding the coding of a particular case (Inter-rater reliability) or the external reliability and agreement of a person at different times of coding in a particular case (Intra-rater reliability) or measures internal reliability.

If the kappa coefficient is below zero, the degree of reliability is poor, between (0.00-0.20) low, (0.21-0.40) relatively poor, (0.41-0.60) moderate, (0.61-0.80) is acceptable and (0.81- 1.00) is almost complete.  $Kappa = \frac{PA0 - PAE}{1 - PAE}$

A value of PA0 indicates the degree of agreement between the two assessors, and a value of PAE indicates the degree of agreement expected (15). In addition to calculating the kappa coefficient, the percentage of agreement was also reported. All analyses were performed using STATA software Version 13.

## Results

In this study, 2015 pathologic reports were examined, (0.57%) 1150 reports did not have topographic, morphology and behavior codes. In other words, 865 (42.9%) reports had topographic code, morphology and behavior codes. Also, only 659 (32.7%) reports had a grade code.

The most reported reports in this study were from the cities of Sari, Amol and Babol and the numbers of these cities' reports were 648 (32.2%), 326 (16.2%) and 250(12.4%).

Governmental centers had 1114(55.3%) reports, of which 496 (57.3%) reports, and they had the codes of topography, morphology and behavior. Based on the allocation of the accurate topographic codes in the 2015 report, the kappa coefficient and the total agreement were 0.266 and 27.70%, respectively, which indicates a lower-than-average accuracy.

Among the cities, Babol had 0.690 (acceptable accuracy), and had the highest kappa coefficient. Of the 865 reported codes, the kappa coefficient and the total agreement, respectively; 0.632 (acceptable) and 64.51% and Tonekabon city with kappa coefficient of 0.837 (almost complete accuracy) is the highest and also Noor, Babol and Amol with kappa coefficients of 0.692, 0.690 and 0.679, were also in the next ranks in terms of accuracy.

To investigate this case, instead of recording the exact code, only the registration of the cancer group was considered. The results showed that among the 2015 reports, the total kappa coefficient was 0.347, which indicates that the accuracy is lower than average. Among the cities, Babol had the highest kappa coefficient with 0.942 (almost complete) (Table 1). According to the standardized classification of cancers, the kappa coefficient in all received reports and reports with topographic codes was 0.346 (relatively good) and 0.906 (almost complete). In the reports with topographic codes, the most reported cancers (percentage of agreement) were related to gastrointestinal cancers 246 (97.6), breast cancers 84 (88.4), female genital cancers 72 (94.7) and skin cancers (table 2).

The results showed that the morphology code accuracy on all received reports and coded reports in pathology centers based on kappa coefficient were 0.331 and 0.834, showed a relatively good and almost complete accuracy, but the accuracy of the behavior and grade codes very low, even on the reported codes (table 3).

The results of the accuracy of the topographic, morphology, behavior and grade codes by center type (public and private), showed that the accuracy is less than average and in all reports, government reporting centers have a higher accuracy in registration. In the reported code, governmental centers almost performed better in registration.

The accuracy of topographic code registration based on cancer grouping and morphology codes is almost complete, in grade code and in behavior code, the accuracy of registration is low (table 4).

**Table 1 - Frequency of reports and accuracy of topographic codes of pathology reports by city**

City	Total number of reports Frequency (percentage)	Topographic, Morphology and Behavior codes Frequency( %)	Based on accurate topographic code registration			Based on cancer grouping registration		
			Kappa coefficient (n=2015)	Kappa coefficient code declaration items (n=865)	Percentage of code declaration agreement (n=865)	Kappa coefficient (n=2015)	Kappa coefficient code declaration items (n=865)	Percentage of code declaration agreement (n=865)
Ramsar	66(3.3)	65(7.5)	0.401	0.407	44.62	0.893	0.913	93.85
Tonekabon	214(10.6)	131(15.1)	0.501	0.837	84.73	0.558	0.955	96.18
Chalous	109(5.4)	23(2.7)	0.127	0.613	65.22	0.156	0.936	95.65
Nowshahr	35(1.7)	14(1.6)	0.160	0.398	42.86	0.281	0.878	92.86
Noor	7(0.3)	4(0.5)	0.378	0.692	75.00	0.475	1.000	100
Mahmoudabad	0(0)	0(0)	-	-	-	-	-	-
Amol	326(16.2)	66(7.6)	0.132	0.679	69.70	0.161	0.903	92.42
Fereydunkenar	23(1.1)	22(2.5)	0.585	0.612	63.64	0.835	0.882	90.91
Babolsar	19(0.9)	18(2.1)	0.288	0.303	33.33	0.576	0.617	77.78
Qaemshahr	155(7.7)	140(16.2)	0.490	0.545	56.43	0.730	0.831	87.14
Sari	648(32.2)	3(0.3)	0.001	0.250	33.33	0.001	0.143	33.33
Neka	30(1.5)	22(2.5)	0.270	0.360	40.91	0.624	0.935	95.45
Behshahr	122(6.1)	107(12.4)	0.502	0.576	59.81	0.822	0.973	98.13
Babol	250(12.4)	250(28.9)	0.690	0.690	70.00	0.942	0.942	95.42
Total Kappa	-	-	0.266	0.632	-	0.347	0.916	-
Total agreement	-	-	27.70%	-	64.51%	39.85%	-	93.45%

\* Cancer groups include: 1- Lips, mouth and throat 2- Digestive organs 3- Respiratory system and organs inside the chest 4- Bones, joints and articular cartilage 5- Blood systems and hematopoietic organs 6- Skin 7- Soft and connective tissue 8- Breast 9- Female genital organ 10- Male genital organ 11- Urinary system 12- Eyes, brain and other parts of central nervous system 13- Thyroid and other endocrine glands 14- Lymph nodes 15- Unknown primary sites 16- ill-defined sites

**Table 2 - Assessing the accuracy of topographic codes given by pathology centers based on the given codes**

ICD-O-3	Topographic category	Total reports received			In reports with topographic codes		
		Coding by experts in cancer registry center N (%)	Coding by experts in pathology centers N (%)	Internal reliability N (%)	Coding by experts in cancer registry center N (%)	Coding by experts in pathology centers N (%)	Internal reliability N (%)
C00- C14	Lips, oral cavity andthroat	32 (1.6)	14 (0.4)	12 (37.5)	10 (1.6)	12 ( 2.0)	10 (100)
C15- C26	Digestive organs	679 (33.7)	355 ( 17.6)	345 (51.0)	252 (41.0)	253 (41.1)	246 (97.6)
C30-C39	Respiratory systemand the chest organs	42 (2.1)	9 (0.4)	9 (21.4)	4 (0.7)	4 (0.7)	4 (100)
C40- C41	Bones, joints andarticular cartilage	16 (0.8)	7 (0.3)	7 (43.8)	8 (1.3)	6 (1.0)	6 (75.0)
C42	Blood systems and hematopoietic organs	89 (4.4)	7 (0.3)	7 (7.9)	2 (0.3)	2( 0.3)	2 (100)
C44	Skin	183 ( 9.1)	81 (3.9)	76 (41.5)	55 (8.9)	57 (9.3)	54 (98.2)
C49	Soft and connective tissue	10 (0.5)	3 (0.1)	3 (30.0)	3 (0.5)	3 (0.5)	3 (100)
C50	Breast	248 (12.3)	118 (5.8)	112 (45.3)	95 (15.4)	88 (14.3)	84 (88.4)
C51-C58	Female genital organ	147 (7.3)	94 (4.6)	83 (56.8)	76 (12.4)	79 (12.8)	72 (94.7)
C60-C63	Male genital organ	139 (6.9)	58 (2.8)	54 (39.1)	45 (7.3)	39 (6.3)	38 (84.4)
C64-C68	Urinary system	190 (9.4)	53 (2.6)	48 (25.5)	33 (5.4)	36 (5.9)	32 (97.0)
C69-C72	Eyes, brain and otherparts of the central nervous system	38 (1.9)	2 (0.1)	2 (5.3)	2 (0.3)	2 (0.3)	2 (100)
C73-C75	Thyroid and otherendocrine glands	29 (1.4)	8( 0.4)	8 (27.6)	2 (0.3)	2 (0.3)	2 (100)
C76	- ill-defined sites	4 (0.2)	5 (0.2)	2 (50.0)	2 (0.3)	4 (0.7)	2 (100)
C77	Lymph nodes	55 (2.7)	15 (0.7)	9 (16.7)	5 (0.8)	10 (1.6)	5 (100)
C80	Unknown primary sites	114 (5.7)	36 (1.8)	22 (19.5)	21 (3.4)	18 (2.9)	8 (38.1)
Kappa coefficient			0.346			0.906	

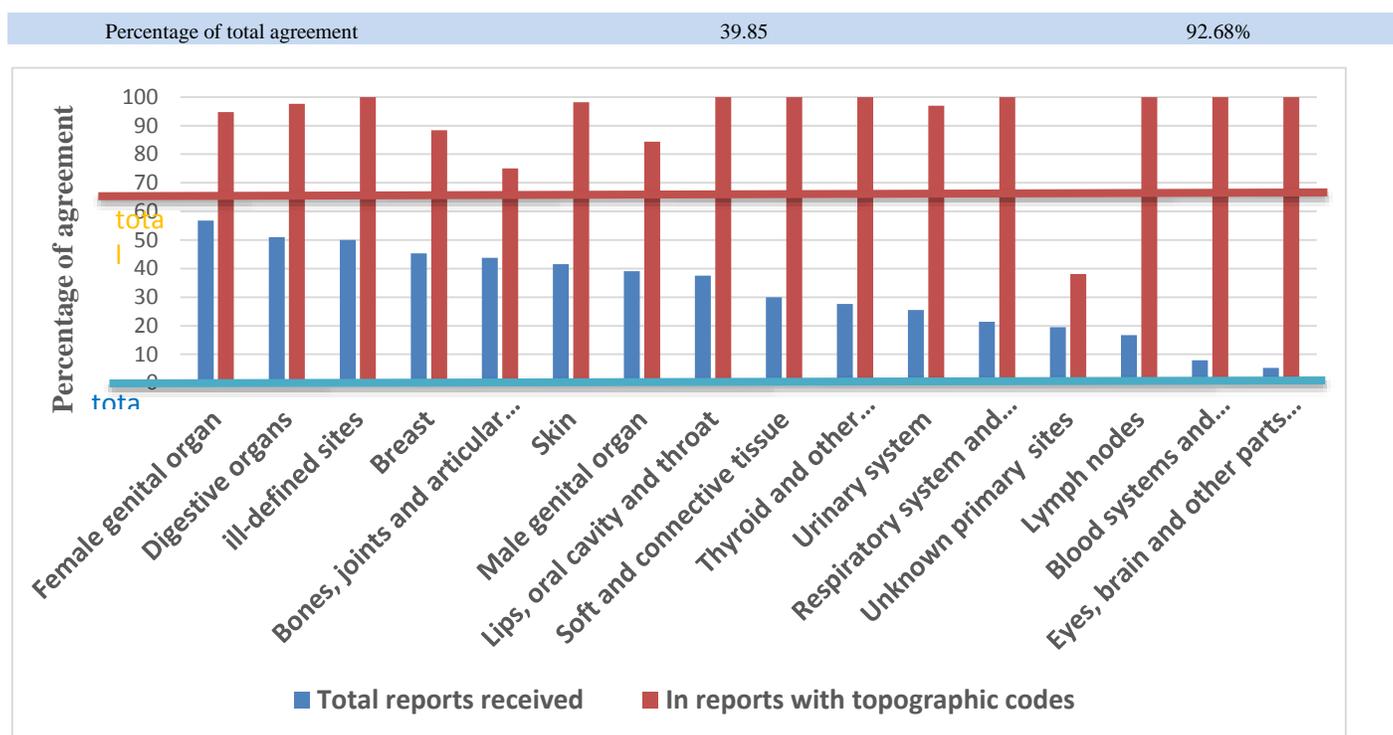


Figure 1- Assessing the percentage of agreement of topographic codes by cancer grouping

Table 3. Accuracy of morphology code, behavior and grade of pathology reports by city

City	Morphology code*			Behavior code**			Grade code***		
	Kappa coefficient (n = 2015)****	Kappa coefficient in reported codes (n = 865)	Percentage agreement coefficient in reported codes (n = 865)	Kappa coefficient in total data (n = 2015)	Kappa coefficient in reported code (n = 865)	Percentage agreement in reported codes (n = 865)	Kappa coefficient in total data (n = 2015)	Kappa coefficient in reported codes (n = 659)	Percentage agreement in reported codes (n = 659)
Ramsar	0.906	0.928	95.38	0.00	0.00	92.31	0.356	0.370	66.15
Tonekabon	0.532	0.913	92.37	0.028	0.664	99.24	0.023	0.632	92.86
Chalous	0.141	0.744	78.26	0.014	0.000	95.65	0.025	0.366	69.23
Nouhshar	0.329	1.000	100	0.000	0.000	0.00	0.082	1.000	100
Nour	0.512	1.000	100	0.000	0.000	0.00	0.097	0.000	0.000
Mahmoud Abad	-	-	-	-	-	-	-	-	-
Amol	0.155	0.841	86.36	0.001	0.010	95.45	0.000	0.000	0.000
Fereydunkenar	0.851	0.895	90.91	0.030	0.023	90.91	0.339	0.564	76.47
Babolsar	0.744	0.792	83.33	0.000	0.000	0.000	0.481	0.544	77.78
Qaeimshahr	0.695	0.779	80.71	0.043	0.052	90.00	0.395	0.492	77.86
Sari	0.002	0.400	66.67	0.000	0.000	0.000	0.000	0.000	0.000
Neka	0.626	0.936	95.45	0.000	0.000	95.45	0.172	0.272	54.55
Behshahr	0.798	0.958	97.20	0.000	0.000	0.000	0.366	0.484	74.29
Babol	0.738	0.738	77.20	0.320	0.320	96.80	0.643	0.643	85.94
Total	0.331	0.834	-	0.027	0.174	-	0.110	0.531	-
Total agreement	%36.92	-	%86.01	%42.98	-	%96.99	%27.34	-	%79.06

\*Morphological codes: carcinoma, lymphoma, leukemia, other specified neoplasms, unspecified neoplasms; \*\* Behavior codes: benign, uncertain, in situ, malignant; \*\*\* Grade codes: grade 0 (no grade), grade 1, grade 2, grade 3, grade 9 (unspecified); \*\*\*\* In total data.

**Table 4. Accuracy of topographic code, morphology, behavior and grade by type of center (public and private)**

Assessment of pathology reports based on	Type of center	Kappa coefficient in total data (n=2015)	Kappa coefficient in reported codes (n=865)	Percentage agreement in reported codes (n=865)
Topographic codes based on accurate code registration	Governmental	0.283	0.647	65.93
	Private	0.244	0.611	62.60
	Total	0.266	0.632	% 64.51
Topographic code based on the grouping of cancers	Governmental	0.367	0.920	93.70
	Private	0.321	0.911	93.11
	Total	0.347	0.916	% 93.45
Morphology code	Governmental	0.341	0.818	84.48
	Private	0.319	0.856	88.08
	Total	0.331	0.834	% 86.01
Behavior Code	Governmental	0.032	0.176	96.57
	Private	0.020	0.170	97.56
	Total	0.027	0.174	% 96.99
Grade Code	Governmental	0.116	0.568	80.90
	Private	0.102	0.484	76.60
	Total	0.110	0.531	% 79.06

## Discussion

The results of the study showed that in the reports received from the pathology centers, based on the accurate topographic codes among all the reports, the accuracy rate was lower than average and among the cities; Babol had the highest accuracy of topographic codes. The total kappa coefficient was acceptable, and the total agreement was above average. The kappa coefficient was relatively good and almost complete in all received reports. In reports with topographic codes, the highest percentage reports of cancers were related to gastrointestinal cancers, breast cancer, female genital cancer and skin cancer. In morphological codes, the accuracy of the codes on the total received reports and the coded reports was relatively good and almost complete, but the accuracy of the behavior and grade codes was very low, even on the reported codes. In general, the reports sent from the pathology centers and also in the reports with the reported codes from the pathology centers, the government centers had a higher accuracy rate and a better performance in registration.

The results of the study showed that based on the registration of the accurate topographic codes among the 2015 reports, the kappa coefficient and the total agreement were 0.266 and 27.70%, which indicates that the accuracy is less than average. Among the cities, Babol with 0.690 (acceptable

accuracy), had the highest kappa coefficient. The results of Beam's research aimed at investigating the specificity of international classification codes of diseases in pulmonary bronchial dysplasia using data in electronic health records and a large insurance database showed that the accuracy of each code was from 82 to 95% (16). The results of Khosravi's study showed that 24% of the causes of death registered in hospitals have a classification error and are in the group of absurd causes of death. This has led to the inefficiency of this information in planning and policymaking(17). Also, the Fall's study entitled Death Certificate in Prostate Cancer showed that the proportion of coded prostate cancer reports due to death from prostate cancer in official death certificates was 3% higher than the records reviewed in the prostate cancer registration program(18).

The present study shows that many reports in Mazandaran University of Medical Sciences are not coded in pathology centers. On the other hand, in Babol University of Medical Sciences, almost all reports are coded in pathology centers, and also if coding is done in pathology centers of these two universities, the accuracy of the codes is generally acceptable. This indicates the accuracy of the coders' coding and regular monitoring of the coding process by the responsible pathologist and the cancer registration team of these

universities, especially in Babol University of Medical Sciences. Also, in the pathology centers of Babol University of Medical Sciences, the latest edition of the ICDO3 coding book is available to all pathologists and experts, and regular coding training and retraining courses have been conducted by cancer registration experts of the University of Medical Sciences to cancer registrars in pathology centers. On the other hand, due to the existence of a comprehensive cancer center and the high volume of referrals to pathology centers in Sari city and the lack of staff time in its pathology centers, coding is done in a very small number and with less accuracy in the centers. Also, in the pathology centers of Babol University of Medical Sciences, a large part of the reports are coded under the supervision of the center pathologist, while in the pathology centers of Mazandaran University of Medical Sciences, a large part of the reports by the center coders coded with the help of existing software and without paying attention to ICDO or without supervision of a pathologist.

According to the standardized classification of cancers, in the reports with topographic codes, the most reports of cancers (percentage of agreement) are related to gastrointestinal cancers 246(97.6), breast cancers 84(88.4), female genital cancers 72(94.7) and skin cancer. In Fink's study that was done to assess the accuracy of cancer mortality statistics based on death certificates of 265863 deaths in which cancer was recorded as the main cause of death and the rate of the overall confirmation was 82.8%. This rate was various in coding the main sites of the disease that led to the death. So that the approval rate in some sites was less than 50% and in some sites were 95% or more (19).

In Johnson's study, assessing the confidence of Qualitative Indicators used in Clinical Cancer Registry Data, concluded that all hospitals currently treating breast cancer patients have the technical capacity to provide data and pathology and surgery data are in high quality. The international classification of diseases, according to the growth of science and the needs of the health community, has always evolved and has played an important role in organizing information and improving treatment processes (20). In Lio's study, assessing the accuracy of disease coding in patients with sarcoma, concluded that the ambiguous definitions of disease could lead to inaccurate coding. As a result, national data sets may not be as comprehensive or useful as expected for studying population-based outcomes for sarcomas (21). The present study confirmed the accuracy of the codes assigned to the reports sent to the cancer registration centers of the

northern universities of the country, the reason was the access to consulting pathologists and oncologists in the cancer registration program of these two universities and holding regular meetings to resolve ambiguities in identifying diagnoses and assigning the correct codes to reports. Also, according to the online registration of reports in the cancer registration system of Mazandaran University of Medical Sciences and daily monitoring of the registration of these reports by the cancer registration experts of this university, many problems in registration are fixed immediately and accurately, which increases accuracy and precision. The coding done by the university's cancer registration experts plays an important role. On the other hand, the lack of complexity in diagnosing the affected site in some recurrent cancers can lead to low errors in their topographic coding.

The results showed that the morphology code accuracy on all received reports and coded reports in pathology centers with kappa coefficient of 0.331 and 0.834 that showed a relatively good and almost complete accuracy but the accuracy of the behavior and grade codes, even on reported codes, was very low. The Wener's study was done to assess the data quality indicators in the cancer registry showed that malignant cancers (excluding non-melanoma skin cancers) diagnosed in the country between 1980 and 2014, and the results showed that the incidence of all tumors also increased with certain fluctuations over time (22). The results of Turner's study showed that the research team stated that death certificates accurately identify the cause of death in men with prostate cancer and that their use in the cancer registry program is trustworthy(23).

In Rashidian's study entitled In-depth exploration of electronic health records stated that these models have the potential to provide coding critiques based on the International Classification of Diseases system that can be used to improve coding accuracy (24). To improve the accuracy of coding the disease, in general, the quality of cancer registration data was acceptable according to the methods presented in this study. The present study also confirmed the accuracy of morphological codes assigned to the reports sent to the cancer registration centers of these medical universities.

The reason for the inaccuracy of the codes of behavior and grade was their non-registration in pathology centers. Overall, the results of this study show that complete coding of diagnoses has become so important that all hospitals were required to register complete codes of diagnoses in the system

(electronic health record), so the correct coding of medical records is essential. Also, the result of the quality of the coded information should be evaluated, and essential steps should be taken to ensure the quality of the coding. Existence of coding errors leads to incorrect classification of diseases and as a result it can cause incorrect statistics. Trust in data whose classification quality is poor, both for care providers and for managers, is a risk in planning and policymaking, educational and research affairs, care and reimbursement.

The results of the accuracy of the topographic, morphology, behavior and grade codes by center types (public and private), showed that the accuracy is less than average and in all cases, governmental centers have a higher accuracy in registration. The study of Jahanbakhsh stated that in all educational and non-educational centers, the average knowledge of coders is less than 50% and indicates a low level of knowledge of coders of the rules and principles of coding (25). The study of Abzari stated that the staff of public and private hospitals had different motivational priorities. Therefore, the managers of these organizations should understand these differences in different employees and try to motivate them (26).

The study of Beasley stated that the staff of health information management in hospital does not only work in this department, but also works in other departments, that each department requires its own specialized skills(27). In general, it can be said that one of the reasons for the inaccuracy of behavior and grade codes is the lack of coded reports in centers, especially the lack of coded reports in private centers. Also, many coders in pathology centers do not have the relevant academic education to perform accurate coding. Therefore, special attention should be paid to holding training courses as well as creating motivational factors for employees in private pathology centers to register these codes more and more accurately and improve the quality of coding. Also, if special attention is paid to the inclusion of code of behavior and grade in the integrated hospital system, this can improve the more accurate recording of these codes in pathology reports and hospital records of patients.

One of the limitations of this research is the incompleteness of collecting all pathology reports in these three years from pathology centers, and also the lack of coding of all pathology reports in pathology centers. The accuracy of the codes given in the pathology centers by cities is in their measurement with the codes given by the coder of the Cancer Registration Center is moderate in topographic codes, is

moderate in morphological codes, is low in behavior codes and is moderate in grade codes. According to the results of the study, if the coding of reports in pathology centers is complete, the available data can be considered reliable for use in epidemiological studies and other research related to cancer. It is recommended to carry out educational interventions to empower and increase the skills of coders, to motivate them to perform more complete coding in pathology centers, as a result of increasing the coding quality of pathology reports. If the coding is done completely and correctly in the pathology centers, they do not need to be re-coded in the cancer registration centers of medical universities, and instead the cancer registration experts can spend more time to monitor and train the coders of the pathology centers.

### Acknowledgments

We would like to thank all the cancer registrars of public and private pathology centers in the cities covered by Mazandaran and Babol Universities of Medical Sciences and the cancer registrars of the Population Based Cancer Registry Centers of Mazandaran and Babol Universities of Medical Sciences with their utmost cooperation in this study.

**Funding:** There was no financial support in the design of the study, data collection, analysis and interpretation of results and writing of the manuscript.

**Conflict of Interest:** The authors declare that there is no conflict of interest.

### Author contribution

R M, MA J and GH M were the principal investigators and designed the study. R M searched literature. R M collected data and prepared data for analysis. MA J and GH M supervised data collection. R M and HA N analyzed data. R M, MA J, GH M and HA N drafted the manuscript and both GH M and MA J supported drafting the manuscript. R M contributed to the implementation of the research. All authors have provided comments and critical revisions to the manuscript.

### References

1. Almomen RK, Kaufman D, Alotaibi H, et al. Applying the ADDIE—analysis, design, development, implementation and evaluation—instructional design model to continuing

- professional development for primary care physicians in Saudi Arabia. *Int J Clin Med* 2016; 7: 538-46.
2. Williams J, Allen L, Wickramasinghe K, et al. A systematic review of associations between non-communicable diseases and socioeconomic status within low-and lower-middle-income countries. *J Global Health* 2018; 8: 20409
  3. Dash S, Shakyawar S, Sharma M. Big data in healthcare: management, analysis and future prospects. *J Big Data* 2019; 6: 1-25.
  4. Mahbubani K, Georgiades F, Goh EL, et al. Clinician-directed improvement in the accuracy of hospital clinical coding. *Future Healthc J* 2018; 5: 47-51.
  5. Cowan AJ, Allen C, Barac A, et al. Global burden of multiple myeloma :a systematic analysis for the global burden of disease study 2016. *JAMA Oncol* 2018; 4: 1221-7.
  6. Jazayeri SB, Saadat S, Ramezani R, Kaviani A. Incidence of primary breast cancer in Iran: Ten-year national cancer registry data report. *Cancer Epidemiol* 2015; 39: 519-27.
  7. Bausat N. Strategies of general hospital of Tenriawaru bone regency in implementation of prospective payment system for delivery care patients in Jamkesda Program. *J Administrasi Rumah Sakit Indonesi* 2015; 1: 97-107. Available at: <https://journal.fkm.ui.ac.id/arsi/article/view/2175/713>
  8. Wang S, Chang X, Li X, et al. Diagnosis code assignment using sparsity-based disease correlation embedding. *IEEE Trans Know Data Eng* 2016; 28: 3191-202.
  9. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; 44: 827-36.
  10. Sheikh A, Cornford T, Barber N, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in “early adopter” hospitals. *BMJ* 2011; 343:d6054
  11. Gjerstorff ML. The Danish cancer registry. *Scand J Public Health* 2011; 39: 42-5.
  12. Stausberg J, Lehmann N, Kaczmarek D, Stein M. Reliability of diagnoses coding with ICD-10. *Int J Med Inform* 2008; 77: 50-7.
  13. Caskey RN, Abutahoun A, Polick A, et al. Transition to International classification of disease version 10, clinical modification: the impact on internal medicine and internal medicine subspecialties. *BMC Health Serv Res* 2018; 18: 328.
  14. Burns EM, Rigby E, Mamidanna R, et al. Systematic review of discharge coding accuracy. *J Public Health* 2012; 34: 138-48.
  15. Stenlund T. Agreement in assessment of prior learning related to higher education: An examination of interrater and intrarater reliability. *Int J Lifelong Educ* 2013; 32: 535-47.
  16. Beam KS, Lee M, Hirst K, Beam A, Parad RB. Specificity of International Classification of Diseases codes for bronchopulmonary dysplasia: an investigation using electronic health record data and a large insurance database. *J Perinatol* 2021; 41: 764-71.
  17. Aghamohamadi S, Khosravi A, Kazemi E, Atefi A. Assessment of the data quality of cause of death registration in Iranian hospitals in 2016. *Iran J Epidemiol* 2020; 16: 90-9. [in Persian]
  18. Fall K, Strömberg F, Rosell J, et al. Reliability of death certificates in prostate cancer patients. *Scand J Urol Nephrol* 2008; 42: 352-7.
  19. Fink AK, German RR, Heron M, et al. Impact of using multiple causes of death codes to compute site-specific, death certificate-based cancer mortality statistics in the United States. *Cancer Epidemiol* 2012; 36: 22-8.
  20. Hartmann-Johnsen OJ, Kåresen R, Schlichting E, Naume B, Nygård JF. Using clinical cancer registry data for estimation of quality indicators: Results from the Norwegian breast cancer registry. *Int J Med Inform* 2019; 125: 102-9.
  21. Lyu HG, Stein LA, Saadat LV, et al. Assessment of the accuracy of disease coding among patients diagnosed with sarcoma. *JAMA Oncol* 2018; 4: 1293-5.
  22. 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.
  23. Turner EL, Metcalfe C, Donovan JL, et al. Contemporary accuracy of death certificates for coding prostate cancer as a cause of death: Is reliance on death certification good enough? A comparison with blinded review by an independent cause of death evaluation committee. *Br J Cancer* 2016; 115: 90-4.
  24. Jahanbakhsh M, Saghaeyannejad IS. Survey of coder's knowledge about coding guidelines in Isfahan hospitals. *Health Inform Manag* 2010; 7: 56-63. [in Persian ]

25. Rashidian S, Hajagos J, Moffitt RA, et al. Deep learning on electronic health records to improve disease coding accuracy. *AMIA Jt Summits Transl Sci Proc* 2019; 2019: 620-9.
26. Abzari M, Shaemi A, Pourmiri M, Azarbaijani K. Study of factors affecting job motivation among staff of private and public hospitals in City of Isfahan. *Health Inform Manag* 2011; 7: 622-8. [in Persian ]
27. Beesley K, McLeod A, Hewitt B, Moczygemba J. Health Information management reimagined: assessing current professional skills and industry demand. *Perspect Health Inf Manag* 2021; 18: 1b.