

Evaluation of Completeness, Comparability, Validity, and Timeliness in Cancer Registries: A Scoping Review

Leila SHOKRIZADEHARANI^a, Zahra BATOOLI^b, Saeedeh HEYDARIAN^a, Reihane SHARIF^{a,1}, Shady GHADERKHANY^a, Maryam TAMEHBIDGOLI^a, Fatemehzahra ATAIEJAHANBEGLOO^a, Daniel HIEBER^c and Peter KUHN^c

^aHealth Information Management Research Center, Department of Health Information Management & Technology, School of Allied Health Professions, Kashan University of Medical Sciences, Kashan, Iran.

^bSocial Determinants of Health (SDH) Research Center, Kashan University of Medical Sciences, Kashan, Iran.

^cInstitute DigiHealth, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany.

ORCID ID: Leila SHOKRIZADEHARANI <https://orcid.org/0000-0002-5586-0580>, Zahra BATOOLI <https://orcid.org/0000-0002-1212-9835>, Saeedeh HEYDARIAN <https://orcid.org/0000-0003-2418-4951>, Reihane SHARIF <https://orcid.org/0000-0002-2081-5093>, Shady GHADERKHANY <https://orcid.org/0000-0002-7805-8187>, Maryam TAMEHBIDGOLI <https://orcid.org/0009-0008-6776-7835>, Fatemehzahra ATAIEJAHANBEGLOO <https://orcid.org/0009-0006-1184-661X>, Daniel HIEBER <https://orcid.org/0000-0002-6278-8759>, Peter KUHN <https://orcid.org/0000-0001-5058-1059>.

Abstract. An essential aspect of cancer registration is data quality. Data quality for Cancer Registries has been reviewed in this paper using four main criteria (comparability, validity, timeliness, and completeness). Medline (via PubMed), Scopus, and Web of Science databases were searched for relevant English articles published from inception until December 2022. Each study was analyzed for its characteristics, measurement method, and data quality features. According to the present study, the majority of articles evaluated the completeness feature, and the fewest evaluated the timeliness feature. A completeness rate of 36% to 99.3% and a timeliness rate of 9% to 98.5% were observed. Standardizing metrics and reporting of data quality is necessary to maintain confidence in the usefulness of cancer registries.

Keywords. Cancer Registries, Data quality, Cancer Documentation

1. Introduction

¹ Corresponding Author: Reihane Sharif. <https://orcid.org/0000-0002-2081-5093> (Email: Fr.sharif@yahoo.com).

Cancer registries are critical to the implementation of national cancer control programs. They are valuable resources for public health, research, evaluation, prevention and early detection, cancer surveillance, and the establishment of control measures. Pathology records (PR) and Hospital discharge records (HDR) are the primary sources for the cancer registry. The level of trust in the data is proportional to its validity, completeness and accuracy of reporting, and timeliness. Cancer population studies require complete, accurate, and reliable baseline data. As a result, the quality of cancer registration data is crucial[1].

A comprehensive literature review on the data quality of cancer registries is needed to provide an overview of these attributes in different registries. This study attempted to take into account evaluating the completeness, timeliness, validity, and comparability of the data in the cancer registries. This study is significant because there has been a lot of interest recently in the use of real-world data for research as well as the creation of new tools and services.

2. Methods

To summarize the body of knowledge on the quality of cancer registry data, a scoping review of the literature was conducted. The Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews are used to report this review.

2.1. Data Sources and Search Strategies

Medline (via PubMed), Scopus, and Web of Science were the electronic databases searched for this study. The search strategy combined Medical Subject Headings (MeSH) terms and keywords in the title and abstract to define three groups: cancer, registry, and data quality. The search period was from inception until December 2022. The details of search strategies are described in Appendix 1.

2.2. Inclusion and Exclusion Criteria

This review includes all English-language papers that discuss at least one of the data quality dimensions (completeness, comparability, validity, and timeliness) based on evaluation guidelines published by the International Agency for Research on Cancer (IARC)[2] and techniques covered by Bray and Parkin[3, 4] in the cancer registration system. Articles were excluded if they were published in a language other than English or involved: (1) Survey and protocol studies, (2) Quality of administrative systems and electronic health records (EHRs), and (3) Registries of clinical trials.

2.3. Data extraction and synthesis

Two researchers independently reviewed the articles' titles, abstracts, and full-texts. Data were extracted from the included studies using a structured form. If necessary, disagreements between researchers were discussed with a third researcher. From each study, the following data were extracted: the study characteristics (including first

author, year published, country, and study design), characteristics of the cancer registry (including name and data evaluated), and measurement method and data quality features[3, 4]. By classifying studies based on data quality dimensions and the ICD-O code, a narrative synthesis was carried out.

3. Results

Figure 1 demonstrates the search results and study selection process using the PRISMA flow diagram. In this review, the completeness feature (n=65) has received the greatest attention, while the timeliness feature (n=11) has received the least. Rates determined were 36%-99.3% for completeness and 9%-98.5% for timeliness. For evaluating completeness, “capture-recapture” methods were used (n=51), and for evaluating comparability, “International standards for classification and coding of neoplasms” were used in most studies (n=10). Reabstracting and recoding was the method for validity (accuracy) that was most frequently employed (n=27). Timeliness was evaluated using the "National guideline" (n=11). Table 1 provides an overview of all utilized measurement approaches. Most of the studies used ICD-O as their classification (n=27). “Block C15-C26 "Malignant neoplasms of the digestive organs" had the largest share with 13 cancer registry subjects.

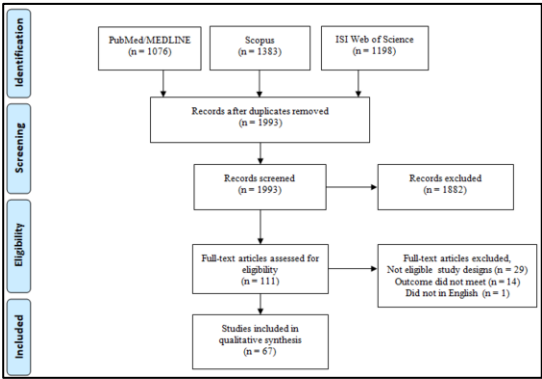


Figure 1. Flow diagram for the selection of studies and literature searches.

Table 1. Data quality measurement approaches.

Dimension	Method	Number of publications
Completeness	Historic data methods • Stability of incidence rates over time • Comparison of incidence rates in different populations • Shape of age-specific curves • Incidence rates of childhood cancers	9
		5
		3
		6
	Mortality: incidence (M:I) ratios	8
	Number of sources/notifications per case	4
	Histological verification of diagnosis	2
	Independent case ascertainment	3
	Capture–recapture methods	51
	Death certificate methods	3
Comparability	International standards for classification and coding of neoplasms (e.g. ICD-10) • DCN/M:I method • The ‘flow’ method	4
		10
		5
	Multiple primaries	2

	Incidental diagnosis	• Screen-detected cancers	1
		• Autopsy diagnosis	1
Validity	Reabstracting and recoding		27
	Histological verification		8
	Death certificate only (DCO)		9
	Missing information		7
	Internal consistency		6
Timeliness	National guideline		11

4. Discussion

According to the findings of this study, data quality in cancer registries was comprehensive and well-functioning, with high levels of coverage and completeness but a low focus on validity, timeliness, and comparability. The study also identifies several areas for improvement. To enable a thorough evaluation and classification of data quality in cancer registries in the absence of standardization framework around data quality for cancer registry, a uniform data model, as well as harmonized categorization criteria and coding rules, are required.

5. Conclusion

In light of the limited comparability and timeliness measurements in studies, as well as the weak measurement method, standardized metrics and data quality reporting are necessary to maintain the general confidence in cancer registries for monitoring cancer care quality and can expand to other disease registries.

References

[1] Wanner M, Matthes KL, Korol D, et al. Indicators of Data Quality at the Cancer Registry Zurich and Zug in Switzerland. *BioMed research international*. 2018;2018:7656197. doi: 10.1155/2018/7656197

[2] Forman D, Bray F, Brewster D, et al. *Cancer Incidence in Five Continents, Volume X*. IARC Scientific Publication No. 164. Lyon, France: International Agency for Research on Cancer. 2014

[3] Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. *European journal of cancer*. 2009;45(5):747-755. doi: 10.1016/j.ejca.2008.11.032

[4] Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. *European journal of cancer*. 2009;45(5):756-764. doi: 10.1016/j.ejca.2008.11.033.

Appendix 1. Database search strategy

(Evaluation OR Accuracy OR Precision OR Conformity OR Completeness OR Timeliness OR Validity OR Uniqueness OR Integrity OR Accessibility OR Relevance OR "Data Quality" OR Representation OR Currency OR Lineage OR "Quality Assessment" OR Redundancy OR Reliability OR Consistency OR "Data Qualities" OR Accuracies OR "Data Management") AND "Cancer Registry"