Minimum Dataset of Lung Cancer Patients: An Initial Step Towards Developing a Web-Based Personal Health Record

Narges Mahmoodi [1], Ali Garavand (iD) [2, *], Mahnaz Samadbeik (iD) [3], Monireh Tahvildarzadeh [4], Ali Asghar Kiani [1]

[1] School of Allied Medical Sciences, Lorestan University of Medical Sciences, Khorramabad, Iran
[2] Department of Health Information Technology, School of Allied Medical Sciences, Lorestan University of Medical Sciences, Khorramabad, Iran
[3] Social Determinants of Health Research Center, Lorestan University of Medical Sciences, Khorramabad, Iran
[4] Medical Education Research Center, Health Management and Safety Promotion Research Institute, Tabriz University of Medical Sciences, Tabriz, Iran

*Corresponding Author: Department of Health Information Technology, School of Allied Medical Sciences, Lorestan University of Medical Sciences, Khorramabad, Iran. Email: virya67@yahoo.com

Abstract

Background: Personal Health Records (PHR), which utilize advanced health information technology tools, play a crucial role in patient self-management and improving the control of chronic diseases such as lung cancer. To optimize the design of these systems, it is essential to determine the necessary data elements.

Objectives: This study aims to identify the minimum dataset required for designing a web-based PHR for lung cancer patients.

Methods: This descriptive, cross-sectional research was conducted in 2023. Initially, a lung cancer dataset was extracted through text analysis. In the next phase, a proposed minimum lung cancer dataset was formulated into a questionnaire containing 18 data groups, including 126 data elements. The dataset underwent expert validation in two phases using the Delphi technique. Data analysis was performed using SPSS version 26, with descriptive statistics employed.

Results: The minimum web-based PHR dataset for lung cancer patients, consisting of 18 data groups (112 data elements), includes demographics, insurance information, emergency contact details, patient symptoms, tumor-related data, physician details, treatment information, patient-reported measurements, personal medical history, history of procedures and surgeries, visits, allergies, family history, medication information, test results, imaging data, dietary information, and educational materials.

Conclusions: Based on the study findings, it is recommended that lung cancer data management encompass not only routine information but also additional dimensions such as allergies, tumor-related information, and dietary details. Collecting comprehensive and complete data can significantly enhance the treatment process and post-treatment follow-up.

Keywords: Lung Neoplasms, Dataset, Personal Health Record

1. Background

Cancer is a large group of diseases that can affect any part of the body, characterized by the abnormal growth and spread of cells, with the ability to metastasize to other areas (1). Lung cancer is one of the most common types of cancer globally and is associated with a high mortality rate (2, 3). Annually, an estimated 1.8 million new cases of lung cancer and nearly 1.5 million lung cancer-related deaths are reported worldwide, with a projected 86% increase in global lung cancer mortality by 2035 compared to 2012 (4). According to the Iranian Ministry of Health, lung cancer is a significant public health concern, ranking as the third leading cause of mortality in the country, following accidents and cardiovascular diseases (5). Lung cancer, as a chronic and life-threatening condition, intensifies anxiety and fear of recurrence, while also negatively affecting the patient's quality of life. Common side effects include hematologic abnormalities, high blood pressure, pneumonia, and fatigue, all of which can significantly

impact both the patient's well-being and the economic outcomes of the disease ([6], [7]).

To improve health outcomes and patient-centered care, active patient engagement in self-management decision-making is essential. Achieving this requires tools that integrate healthcare data, clinical evidence, and facilitate communication between care providers. One such tool is the Personal Health Record (PHR), which is considered a solution to this challenge ([8]). The use of PHR by patients can lead to cost savings and improve public health outcomes, especially for chronic diseases ([9]). In recent decades, PHRs have been deployed to enhance healthcare services for patients with chronic conditions ([10]).

Advancements in PHR technology have enabled the development of new self-management tools, data exchange capabilities, and web-based sharing, allowing patients to take a more active role in their self-care. Given the increasing survival rates for cancer patients, the complexities of cancer care, and the demands of oncology management, PHRs play a critical role in achieving these goals ([11]). Since the data collected by PHRs must support effective self-management and improve patient health, defining a Minimum Data Set (MDS) for PHRs is essential ([12]).

An MDS serves as a standardized tool for collecting healthcare data, providing precise and consistent access to disease-related information. It enhances communication between healthcare providers, streamlines decision-making for managers ([13]), and ensures consistency in data definitions, allowing for data comparison at both national and international levels ([12]).

Many studies have been conducted to identify and define MDS for various diseases ([12-14]). However, no study in Iran has addressed the development of an MDS for lung cancer.

2. Objectives

Thus, this study aimed to determine the MDS for designing a PHR specifically for lung cancer patients.

3. Methods

The current study was a descriptive cross-sectional study aimed at identifying essential data elements for designing a web-based PHR for lung cancer patients. The study was conducted in two phases. First, a review was carried out, followed by the Delphi technique to determine and confirm the MDS for lung cancer.

3.1. Data Sources

To collect data, a review was conducted using databases such as PubMed, Scopus, Web of Science, and relevant websites and textbooks. Published articles up until 2023 were identified and included in the study. Additionally, medical records of lung cancer patients were examined to supplement the data items and design based on the MDS needs. The search strategy employed a combination of keywords, including "personal health record," "electronic personal health record," "ePHR," "personal electronic health record," "web-based personal health record," "lung cancer," and "lung adenocarcinoma."

The inclusion criteria for this study were: (1) all English-language research articles that provided sources for extracting data items, characteristics, and required functions to design a web-based PHR for lung cancer patients, and (2) research articles, regardless of type or timing, to avoid overlooking potential sources. Articles published in languages other than English were excluded if the full text was not available in English. Additionally, short communications, letters to the editor, and review articles were excluded from the study.

3.2. Data Collection

Subsequently, the proposed MDS for lung cancer was developed in the form of a researcher-made questionnaire. This questionnaire was designed to collect opinions on informational elements both electronically and in paper form, to be reviewed by the relevant stakeholders as outlined in the sampling process.

Data collection was conducted using this questionnaire, which contained the MDS for the PHR of lung cancer patients. It was organized into 18 data groups with a total of 126 questions. The questionnaire presented participants with five-point Likert scale questions (1 = lowest score, 5 = highest score), asking them to indicate their level of agreement on the inclusion of each data element in the PHR for lung cancer. Additionally, an open-ended section was included, allowing participants to add any specific suggestions or considerations for further review in the next phase of feedback.

3.3. Sampling

Sampling was conducted based on the study by Azizi et al. ([15]), utilizing non-random convenience sampling methods. A total of 21 individuals were selected for the study, including 3 oncologists, 4 pulmonologists, 7 health information management experts, 2 medical informatics experts, 1 expert nurse with a Ph.D., 1 nurse

with a master's degree, and 3 nurses with bachelor's degrees (all working in the lung ward). Of the 21 distributed questionnaires, 19 were completed and returned.

3.4. Data Analysis

In the review phase, the selected articles were examined, and relevant items and information were extracted using a data extraction form. The content of the materials under review was analyzed using the content analysis technique. For the Delphi study phase, data analysis was conducted using SPSS version 26, employing descriptive statistics such as mean and percentage agreement. The decision to approve or reject each data element was based on the percentage of agreement among the experts. Data elements with an average score of over 75% were accepted, those with a score below 50% were removed, and elements scoring between 50% and 75% were included in the second phase of the Delphi study for further evaluation.

4. Results

To determine the MDS, data elements were identified through a literature review and medical record analysis. These elements were extracted and used to design the preliminary questionnaire, which was then presented to experts in the field using the Delphi method. Of the 21 people approached, 19 (90.4%) agreed to participate in the study. The demographic details of the participants are displayed in Table 1. According to the table, 57.9% of the participants were female. The average age of the participants was 40.1 years, and the average work experience was 11.78 years.

In the first phase, the questionnaire comprised 18 data groups and 126 questions. During the initial round of the Delphi technique, 104 data elements were approved, while 8 data elements did not receive consensus from the participants, and 14 data elements advanced to the second round for further evaluation. Additionally, 6 new data elements were suggested by the participants. In the second round of the Delphi process, 8 out of the 20 data elements were approved, while 12 were removed. Ultimately, across both Delphi phases, a total of 112 data elements, organized into 18 data groups, received final approval for inclusion in the MDS for the web-based PHR for lung cancer patients, as shown in Table 2.

The minimum dataset for the web-based PHR of lung cancer patients comprises 112 data elements, organized into 18 data groups. In the first round of the Delphi process, certain data elements, including email, time of

chemotherapy, time of radiotherapy, radiation dose, method of biopsy, required protein, and required energy, were removed due to lack of consensus among the study participants.

5. Discussion

The MDS encompasses healthcare and treatment effectiveness data, including both general information and specific disease-related data (such as disease assessment, staging, progression, and risk factors). This dataset serves as a foundation for achieving key effectiveness indicators and ensures access to clear and accurate health data on diseases (16, 17). Developing a comprehensive, standardized MDS is essential for delivering more efficient and effective services to lung cancer patients through PHR systems.

The current study classifies data in the MDS into general and specific categories, covering both general data (such as demographic details, insurance information, and emergency contacts) and disease-specific data (such as personal medical history, previous surgeries, allergies, family history, medications, and laboratory test results). Previous studies (18-22) have underscored the importance of including: Demographic information, test results, and medications in the health records of lung cancer patients. These elements are crucial for the effective use of PHRs by lung cancer patients, as reflected in the findings.

Additionally, this study highlights the importance of including information on allergies in lung cancer patients' health records, as well as data on diagnostic imaging and other comorbid conditions. Demographic data, family history, healthcare providers, consulting physicians, allergies, surgeries, chronic conditions, medications, and imaging are key components for designing and implementing a PHR system (19). Given the necessity of these data dimensions, it is recommended that developers of lung cancer information systems, in addition to identity and insurance information, prioritize these essential items, which play a vital role in patient treatment and follow-up.

The studies by Yuan et al. (23) and Griffith et al. (24) also utilized data elements such as demographic characteristics, smoking status, tumor information, treatment details (chemotherapy, radiotherapy, etc.), and personal medical history in the electronic health records of lung cancer patients. This highlights the importance and necessity of incorporating these informational elements in the PHR of lung cancer patients, which aligns with the findings of the current study and has been carefully considered.

Table 1. Demographic Information of Participants in the Delphi Phase [a]

Variables	Values
Age	40.1 ± 5.98
Work experience	11.78 ± 7.56
Gender	
Female	11 (57.9)
Male	8 (42.1)
Education level	
BSc	3 (15.8)
MSc	1 (5.3)
Ph.D.	9 (47.4)
MD.	6 (31.6)
Field of study	
Lung specialist	3 (15.8)
Oncologist	3 (15.8)
Health information management	6 (31.6)
Medical informatics	2 (10.5)
Nurse	5 (26.4)

[a] Values are expressed as No. (%) or mean ± SD.

In domestic research, Jeddi et al. identified essential data elements for designing PHR for patients undergoing chronic dialysis, including demographic information, insurance details, emergency contact information, physician information, individual measurements, medical history, surgical methods and procedures, visit history, allergies, vaccinations, family history, medications, laboratory results, dietary regimen, and patient education materials (12). These findings support the inclusion of these data elements in PHRs, as reaffirmed by the current study.

Similarly, Maghsoudi et al., in developing a PHR system for patients with thyroid dysfunction, identified six key data elements: Demographic data, medical history, general examinations, laboratory results, treatment plan, and diagnostic information (25). These elements have also been incorporated into the current study, demonstrating their importance in PHR design for patients with thyroid dysfunction.

The analysis of previous studies, combined with the findings of the current study, underscores the necessity and significance of the 18 data groups derived from this research in the design of health information systems, especially PHR systems. These data groups are crucial in providing comprehensive and effective care, supporting personalized healthcare services tailored to individual patient needs.

5.1. Conclusions

The development of self-management software for various types of cancer, especially lung cancer, which requires specialized care, can significantly reduce the impact of this disease. Utilizing a comprehensive and well-structured MDS tailored to the internal needs of the country can enhance the effectiveness of these software applications. By customizing the software to address the specific challenges faced by patients and healthcare providers, these tools can substantially improve the management and monitoring of the condition, ultimately leading to better patient outcomes and quality of life.

The findings of the current study represent an important first step toward the development of a PHR for lung cancer patients by introducing a comprehensive MDS that can be utilized in future research. This MDS serves as a foundation for the creation of personalized healthcare solutions specifically designed to meet the needs of lung cancer patients. It offers a framework that can contribute to improved care, treatment outcomes, and overall quality

Table 2. Minimum Data Set of Personal Health Records for Patients with Lung Cancer

Row	Data Categories	Data Element
1	Demographics details	Name, father's name, national ID, date of birth, gender, place of birth, marital status, mobile phone number, home address, occupation, smoking or secondhand smoke exposure status, substance abuse/alcohol addiction status, daily tobacco consumption amount, duration of tobacco/drug/alcohol use, number of tobacco users in the family, number of drug/alcohol users in the family, number of friends who use tobacco/alcohol.
2	Insurance information	Insurance organization, type of insurance.
3	Emergency contact information	Name, phone, address, relationship, mobile phone.
4	Information related to the patient's symptoms	Respiratory and non-respiratory symptoms.
5	Information related to the tumor	Date of diagnosis, date under observation, tumor pathology, T (Tumor size), N (involvement of lymph nodes), M (metastasis to other organs), staging based on M, N, T, involved lobe, side involved, disease stage post-surgical operation.
6	Physicians' information	Doctor's first and last name, specialty, phone number, address
7	Treatment information	Chemotherapy start date, chemotherapy drugs, chemotherapy side effects, radiation therapy start date, radiation therapy side effects, type of surgical operation, surgical operation side effects.
8	Medical history	History of heart disease, history of kidney disease, history of diabetes, history of other respiratory diseases, and other medical conditions.
9	Surgical procedures and interventions	Name of procedure or surgical operation, date of procedure, body location, treatment center where it was performed, side effects, outcome.
10	Visit records	Date of visit, name of physician, reason for visit, recommendations.
11	Allergies	Type of allergy, date of the first encounter, allergy symptoms, allergen, has treatment been administered, Has the allergy treatment been stopped?
12	Self-measured data	Date of measurement, weight, blood sugar level, time of blood sugar measurement, systolic blood pressure, diastolic blood pressure.
13	Family history	Relationship to the patient, names of significant illnesses.
14	Medications	Drug name, dosage, dosage form, reason for prescription, administration time, start date of use, stop date of use, number and frequency of drug use, side effects, prescriber, interactions with other drugs or food.
15	Laboratory tests	Date of examination, name of test, result, recommendations, pulmonary test function, blood pH level, analysis of pleuritic fluid
16	Imaging studies	Date of imaging, type of imaging, with or without contrast injection, imaging result, recommendations.
17	Dietary regimen	Body Mass Index (BMI), malnutrition severity, type of dietary regimen, dietary interactions with medications taken, and dietary recommendations.
18	Educational materials for the patient	Patient's familiarity with disease symptoms and signs, treatment methods for lung cancer patients, exercise and physical activity methods, principles of using respiratory assistance devices (such as BiPAP), and tracheostomy care principles.

of life for individuals facing this complex and challenging disease.

Based on the results of the study, it is recommended that in addition to routine data, various other dimensions such as allergies, tumor information, and dietary regimen should also be collected in the management of lung cancer data. Collecting comprehensive and complete data will facilitate the treatment process, assist with post-treatment follow-up, and ultimately lead to better patient care and outcomes in the management of lung cancer.

Footnotes

References

1. World Health Organization. *Cancer*. 2023. Available from: https://www.who.int/health-topics/cancer#tab=tab_1.

2. Sharma P, Mehta M, Dhanjal DS, Kaur S, Gupta G, Singh H, et al. Emerging trends in the novel drug delivery approaches for the treatment of lung cancer. *Chem Biol Interact*. 2019;**309**:108720. [PubMed ID: 31226287]. https://doi.org/10.1016/j.cbi.2019.06.033.

3. Barta JA, Powell CA, Wisnivesky JP. Global Epidemiology of Lung Cancer. *Ann Glob Health*. 2019;**85**(1). [PubMed ID: 30741509]. [PubMed Central ID: PMC6724220]. https://doi.org/10.5334/aogh.2419.

4. Zubi ZS, Saad RA. Improves Treatment Programs of Lung Cancer Using Data Mining Techniques. *Journal of Software Engineering and Applications*. 2014;7(2):69-77. https://doi.org/10.4236/jsea.2014.72008.

5. Rajan JR, Prakash JJ. Early diagnosis of lung cancer using a mining tool. *International Journal of Emerging Trends and Technology in Computer Science*. 2013.

6. Rowntree RA, Hosseinzadeh H. Lung Cancer and Self-Management Interventions: A Systematic Review of Randomised Controlled Trials. *Int J Environ Res Public Health*. 2022;**19**(1). [PubMed ID: 35010796]. [PubMed Central ID: PMC8744740]. https://doi.org/10.3390/ijerph19010536.

7. Elsayed AM, Ahmed OA, Mohamed AH, Mahrous FM, Sherif DEM. Effect of self-care guidelines on symptoms burden for patients with lung cancer undergoing chemotherapy. *Evidence-Based Nursing Research*. 2021;**3**(4):14. https://doi.org/10.47104/ebnrojs3.v3i4.214.

8. Davis S, Roudsari A, Courtney KL. Designing Personal Health Record Technology for Shared Decision Making. *Stud Health Technol Inform*. 2017;**234**:75-80. [PubMed ID: 28186019].

9. Andrikopoulou E, Scott P. Personal Health Records an Approach to Answer: What Works for Whom in What Circumstances? *Stud Health Technol Inform*. 2022;**294**:725-9. [PubMed ID: 35612192]. https://doi.org/10.3233/SHTI220572.

10. Aslani N, Ahmadi M, Samadbeik M. A systematic review of the attributes of electronic personal health Records for Patients with multiple sclerosis. *Health and Technology*. 2019;**10**(3):587-99. https://doi.org/10.1007/s12553-019-00387-4.

11. Haggstrom DA, Carr T. Uses of Personal Health Records for Communication Among Colorectal Cancer Survivors, Caregivers, and Providers: Interview and Observational Study in a Human-Computer Interaction Laboratory. *JMIR Hum Factors*. 2022;**9**(1). e16447. [PubMed ID: 35076406]. [PubMed Central ID: PMC8826153]. https://doi.org/10.2196/16447.

12. Jeddi FR, Nabovati E, Rezayi S, Saeedi S, Amirazodi S. Identifying a minimum data set as a necessity to design a web-based personal health record for patients under chronic dialysis. *J Family Med Prim Care*. 2022;**11**(3):969-75. [PubMed ID: 35495826]. [PubMed Central ID: PMC9051723]. https://doi.org/10.4103/jfmpc.jfmpc_978_21.

13. Bahador F, Javanmard Z, Sabahi A, Salehi F, Shahabi S, Fateh Bahari R. [Determining the Minimum Data Set for Electronic Health Records of Tuberculosis Patients]. *J Health Biomed Informatics*. 2024;**10**(4):380-5. FA. https://doi.org/10.34172/jhbmi.2024.06.

14. Ghazi Saeedi M, Esmaeili Javid GR, Mohammadzadeh N, Asadallah Khan vali H. [Determining the Minimum Data Set of Diabetic Foot Ulcer Patients' Electronic Medical Records]. *Payavard Salamat*. 2021;**14**(5):443-53. FA.

15. Azizi A, Aboutorabi R, Mazloum-Khorasani Z, Afzal-Aghaea M, Tabesh H, Tara M. Evaluating the Effect of Web-Based Iranian Diabetic Personal Health Record App on Self-Care Status and Clinical Indicators: Randomized Controlled Trial. *JMIR Med Inform*. 2016;**4**(4). e32. [PubMed ID: 27769953]. [PubMed Central ID: PMC5097176]. https://doi.org/10.2196/medinform.6433.

16. Hosseini A, Moghaddasi H, Jahanbakhsh M. [Designing minimum data sets of diabetes mellitus: Basis of effectiveness indicators of diabetes management]. *Health Info Manage*. 2010;**7**(3). FA.

17. Sadoughi F, Safdari R, Meraji M, Ramzan Ghorbani N, Ghazisaeedi M. [Designing a minimum data set for national cancer registry in Iran]. *Nurs Mid J*. 2013;**11**(8):0. FA.

18. Krist AH, Peele E, Woolf SH, Rothemich SF, Loomis JF, Longo DR, et al. Designing a patient-centered personal health record to promote preventive care. *BMC Med Inform Decis Mak*. 2011;**11**:73. [PubMed ID: 22115059]. [PubMed Central ID: PMC3250934]. https://doi.org/10.1186/1472-6947-11-73.

19. Kim MI, Johnson KB. Personal health records: evaluation of functionality and utility. *J Am Med Inform Assoc*. 2002;**9**(2):171-80. [PubMed ID: 11861632]. [PubMed Central ID: PMC344574]. https://doi.org/10.1197/jamia.m0978.

20. Segall N, Saville JG, L'Engle P, Carlson B, Wright MC, Schulman K, et al. Usability evaluation of a personal health record. *AMIA Annu Symp Proc*. 2011;**2011**:1233-42. [PubMed ID: 22195184]. [PubMed Central ID: PMC3243224].

21. Moeil Tabaghdehi K, Ghazisaeedi M, Shahmoradi L, Karami H. Designing and creating personal electronic health records for thalassemia major patients. *Payavard Salamat*. 2018;**11**(5):567-77.

22. Kyazze M, Wesson J, Naude K. The design and implementation of a ubiquitous personal health record system for South Africa. *Stud Health Technol Inform*. 2014;**206**:29-41. [PubMed ID: 25365669].

23. Yuan Q, Cai T, Hong C, Du M, Johnson BE, Lanuti M, et al. Performance of a machine learning algorithm using electronic health record data to identify and estimate survival in a longitudinal cohort of patients with lung cancer. *JAMA Network Open*. 2021;**4**(7):e2114723. https://doi.org/10.1001/jamanetworkopen.2021.14723.

24. Griffith SD, Tucker M, Bowser B, Calkins G, Chang CJ, Guardino E, et al. Generating Real-World Tumor Burden Endpoints from Electronic Health Record Data: Comparison of RECIST, Radiology-Anchored, and Clinician-Anchored Approaches for Abstracting Real-World Progression in Non-Small Cell Lung Cancer. *Adv Ther*. 2019;**36**(8):2122-36. [PubMed ID: 31140124]. [PubMed Central ID: PMC6822856]. https://doi.org/10.1007/s12325-019-00970-1.

25. Maghsoudi B, Langarizadeh M, Esmaeili R, Movahedi Rad M. [Providing a Personal Health Record System for Patients with Hypothyroid Disease]. *J Health Biomedical Informatics*. 2022;**9**(2):50-7. FA.