EUROPEAN ACADEMIC RESEARCH
Vol. I, Issue 12/ March 2014

ISSN 2286-4822
www.euacademic.org

Impact Factor: 3.1 (UIF)
DRJI Value: 5.9 (B+)

Collection and Analysis of the Lymphoma Cancer Related Data and its Systematization by Developing A Database

MUHAMMAD SARFARAZ IQBAL
HAFIZ MUHAMMAD AHMAD[1]
MUHAMMAD USMAN GHANI
MUHAMMAD AMJAD ALI
Department of Bioinformatics and Biotechnology
Government College University, Faisalabad
Pakistan

Abstract:

A study was carried out at Punjab Institute of Nuclear Medicine (PINUM) cancer Hospital, Faisalabad during the period of June 2010 to June 2011 to access the percentage of lymphoma patient on the bases of age, gender, body mass index and economical status to make the database that will allow scientists and clinicians to search cancer related data and information across the spectrum. This database provides the help for management of lymphoma cancer data and replaces the previous paper based medical record with computer-based record system. For making data base 690 files of already diagnosed malignant patients were studied. Out of 690 patients, 110 (16%) patients were diagnosed with lymphoma cancer. Out of 110 lymphoma cases, 36 (33%) cases of Hodgkin Lymphoma and 74 (67%) cases of Non Hodgkin Lymphoma were observed. Gender based study showed that among 110 patients, 81 (74%) were male and 29(26%) were female. Patient age ranged from 7 years to 85 years, mean age being 46 years.

Key words: Database tools, Cancer, lymphoma cancer, bioinformatics

[1] Corresponding author: hafizahmad90@yahoo.com

Introduction:

The word cancer is derived from Greek word, "carcinos", which means crab, generally meaning a deeply penetrating type of ulceration and was credited by the Greek physician Hippocrates. Lymphomas are a heterogeneous group of cancers that are characterized by abnormal growth of tissue in the lymphatic system. These disorders originate from B-lymphocytes, T-lymphocytes, and natural killer (NK) cells. The initial recognition of lymphomas as specific clinical entities is often ascribed to Thomas Hodgkin's observations in the (Hodgkin 1832). Reed and Sternberg independently described the cell that bears their name and is characteristic of the entity called Hodgkin lymphoma (Reed 1902). Lymphoma cancer is the fifth most frequent malignancy among both genders of all age groups in the world.

As far as Pakistan is concerned unfortunately the incidence of lymphoma is reported highest in any South-Central Asian country. The exact death ratio and number of new cases of lymphoma is difficult to measure at the national and community levels due to lack of cancer registration schemes. Bioinformatics as the field at the crossroads of biology and computational engineering responsible for the storage, distribution, extraction analysis and presentation of complex biological information in understandable forms, gives solutions for developing hospital base and national level cancer registries.

As the incidences of lymphoma are increasing tremendously, therefore large amount of clinical data about lymphoma is available. This data will be managed, systematize and utilized in developing a deeper understanding of risk factors, prognostic factors and in determining the recent trends of lymphoma. The database does not only provide assistance to the oncologists but it also provided a base for the development of a risk assessment tool. Today, the medical community is in a

state of transition from a situation dominated by the paper medical record to a future situation where computer-based patient record systems will be available to store all patient data (Lucas *et al.* 2000). In Pakistan the incidences of lymphoma is growing rapidly, the exact death ratio and number of new cases of lymphoma is difficult to measure, because in Pakistani hospitals there is no computational method used for storing and retrieving patient information, the data available is in the raw form and in file format. Computational tools and databases are essential to the management and identification of patterns among database elements that reflect biological systems (Buehler and Rashidi 2005). The aim and objective of the present study was to collect and analyze the lymphoma cancer related data in order to report recent trends regarding prognostic factors as well as risk factors and systematize the data by developing a database.

Literature Review:

Non- Hodgkin's Lymphoma (NHL) is the most common childhood cancer in the population of south Karachi. In Pakistan, increase in incidences of the lymphoma affects all age groups in both genders. Pakistani males are six times more at risk than female of developing Hodgkin lymphoma and men are three times more at risk than female of developing Non Hodgkin lymphoma. The variation in the incidence of lymphoma is strongly reflected in the epidemiological pattern of cancers in Pakistan. **(**Bhurgri et al. 2005) Frequency of the small B-cell Non Hodgkin lymphoma (NHL) is very low in Pakistani population as compared to the western literature. Majority of the small B-cell NHL in Pakistani population were nodal. Small B- cell NHL was more common in Pakistani males, with male to female ratio of 2.1. Small B-cell Non Hodgkin Lymphoma is relatively uncommon in Pakistan. The second commonest small B-cell non-Hodgkin's lymphoma in

their observation was Follicular lymphoma and it was 2.6% of all NHL. They further reported that biological studies needed to determine the cause of a higher component of aggressive NHL in Pakistani population in comparison to the low grade NHL. **(**Aftab, Bhurgri, and Pervez, 2006**).** Oh et al. (2006) conducted an experiment to study about four thousand two hundred forty-six cancer patients using prostate cancer clinical research information system. Mean age of patients was 62 years, and 89% were white. Seventy-one percent of patients presented at diagnosis with T1 or T2 disease, and 78% had biopsy Gleason scores of <or=7, 8-10 in 18%. Median prostate-specific antigen level at diagnosis was 7 ng/mL, and 77% of patients presented with increased prostate-specific antigen as a trigger symptom. Sixty-four percent of patients presented to our clinic having had no previous treatment for prostate cancer. The majority of approached patients provided consent for collection of clinical data, blood, and tissue. Quality control assessments demonstrated high levels of concordance among data entry personnel. Huang et al 2009 developed a cancer Biomedical Informatics Grid™ (caBIG™) Silver level compliant lymphoma database, called the Lymphoma Enterprise Architecture Data-system™ (LEAD™), which integrated the pathology, pharmacy, laboratory, cancer registry, clinical trials, and clinical data from institutional databases. They also utilized the Cancer Common Ontological Representation. They reported that the data elements and structures within LEAD™ could be used to manage clinical research data from phase-1 clinical trials, cohort studies, and registry data from the Surveillance Epidemiology and End Results database. Their work provided a clear example of how semantic technologies from caBIG™ can be applied to support a wide range of clinical and research tasks, and integrate data from disparate systems into a single architecture.

Methodology:

The current work deals with the design, development of a computerized hospital based cancer registry database and cancer risk assessment tool to assist the healthcare providers with a simplified solution in easy documentation and instant access of health information of patients with cancer for care better planning, research, education, and national cancer registry reporting to governments and funders. This cancer registry database will be a vital tool for programmatic and administrative planning, new research and monitoring of lymphoma cancer patient outcomes. This descriptive study was carried out at Punjab Institute of Nuclear Medicine(PINUM) cancer Hospital Faisalabad department including 300 cases of Lymphoma diagnosed during a period of one year (June 2010-June 2011).We gathered the known population-based sample of lymphoma cases and classified according to the current REAL/WHO classification of hematopoietic and lymphoid tumors. For the purpose of data collection a Performa was designed by taking in consideration the possible parameters and filled by extracting information from the patient's record files. Missing information was gathered by telephonically and by visiting patients. Data related to possible risk factors like family history, medical history and life style related information. Prognostic factor like receptor status, histopathology, grade of tumor etc was also collected. Information on host characteristics, history of cancer and risk factors associated with lymphoma was collected during identical standardized telephone interviews and take-home questionnaires for patients. We collect and retain the information necessary for the designing and development of the lymphoma cancer database and risk assessment tool. The data was collected according to maintain the data standards and coding instructions. The present study covered the population of Pakistan, including all residents of all age groups. Within the

study base and time period, 690 patients files were analyzed, of all those diagnosed with malignant tumors or cancer. Out of 690 cases 115 cases of lymphoma were found. We selected all cases of lymphoma and uniformly classified according to the World Health Organization classification system. To understand more fully the conditions associated with lymphoma, we asked questions about the demographic characteristics of the respondents and their families, health conditions, family well-being, and current access to health care. We also recorded host characteristics, including age, sex, birth order (first, second, or third or higher), address, current height, normal adult weight, family background, smoking habits, sun exposure, medical history, family medical history, education, occupation, childhood environment, educational level, job history, occupation lasting at least 1 year (ever or never), occupation involving exposure to pesticides (yes or no) or to organic solvents. All study participants completed a telephone interview including questions on and other possible risk factors for lymphoma. After designing the database architecture we build a database schema by listing the table names, field names within the tables, field size, field type, key fields, what information the field will contain and how it relates to other fields to ensure the DB reliability and enhanced performance. We defined data attributes, display attributes, valid values, validation rules, data entry rules, and other documentation for each data element in this phase. We designed a lymphoma cancer relational database by using "Structured Query Language "MySQL which is the most common database language and is the most popular "open source" database in the World. For this purpose we used Apache Server to work with MySQL files which is available in WAMP server. For designing a web base lymphoma cancer database we select WAMP2.0c a mini-server that can run on almost any Windows Operating System. Wamp server is abbreviated for Windows-Apache-MySQL-PHP server and refers to a set of free (open source)

applications, combined with Microsoft Windows, which are commonly used in Web server environments. The WAMP2.0c stack contains the four key elements of a Web server: an operating system, database, Web server and Web scripting software. The combined usage of these four programs is called a server stack. In the server stack, Microsoft Windows is the operating system (OS), Apache is the Web server, and MySQL handles the database components, while PHP, Python, or PERL represents the dynamic scripting languages. The WAMP2.0c includes the following programming and database management applications:

Statistical Analyses

Relative risk (RR): The relative risk was calculated by using the formula:

$$RR=P \text{ exposed}/P \text{ non-exposed}$$

Population attributable risk (PAR): The PAR was calculated by subtracting the incidence in the unexposed (Iu) from the incidence in total population (exposed and unexposed) (Ip)

$$PAR = Ip - Iu$$

Development of Lymphoma Risk Assessment Tool:

In order to drastically minimize the risk of tool failure, we approached the application development projects in the following sequence, Identify logic and entities, designing of a functional specification and project plan, Prototype development, development of support scheme for the tool, resting, support and stability.

Table Name	Description and fields
Database suspense system	This database contains a suspense system that provides a temporary storage for potential cases. A new patient data is entered into this system. The suspense system typically includes the following: 1. Patient name 2. Date of birth 3. Sex 4. Patient identifier - 5. medical record/PINUM number 6. Date of first contact /diagnosis date 7. Primary site 8. Family history of cancer 9. Prognostic factors 10. Treatment plan 11. Stage of the cancer
Patent Index	The patient index is an alphabetical list of each patient entered into the registry since the reference date. The typical index includes Patient ID, Patient name and gender. activity category identifier
Abstract	This table provide summary of the patient information and give overview about the stage and sort of cancer.
User	User information is stored on this table. It consists of use name, email address, login ID, password and user access level.
Access log	When a user enters the system, user login data are saved in this table. It contains user identifier number, login date and time, IP(Internet Protocol) address, and etc.

Results and discussion:

According to the recent studies the prevalence of lymphoma in Pakistan is alarming and particularly at PINUM Cancer Hospital ratio of lymphoma cancer patients as compared to other sort of cancers are much higher. Where, it accounts for 16% of all cancers at PINUM Cancer Hospital (Fig.1). This data can be managed, systematize and utilized in developing a deeper understanding of risk factors, prognostic factors and in determining the recent trends of lymphoma. All this has been achieved by developing lymphoma Cancer Database. A total 690

record files of already diagnosed malignant patients were studied .Out of 690 patients, 110 (16%) patients were diagnosed with lymphoma cancer. found and studied. Of 110 lymphoma cases, 36 (33%) cases of Hodgkin disease and 74 (67%) cases of Non Hodgkin Lymphoma were observed. Of 110 patients, 81 (74%) were male and 29(26%) were female. Patient age ranged from 7 years to 85 years, mean age being 46 years. The top 10 malignancies that observed at PINUM hospital Faisalabad are presented in fig. 1. The presenting characteristics of 110 patients with lymphoma who were included in analysis are shown in the table. The results agree with Bhurgri et al., 2005 who reported similar results for lymphoma in Karachi.

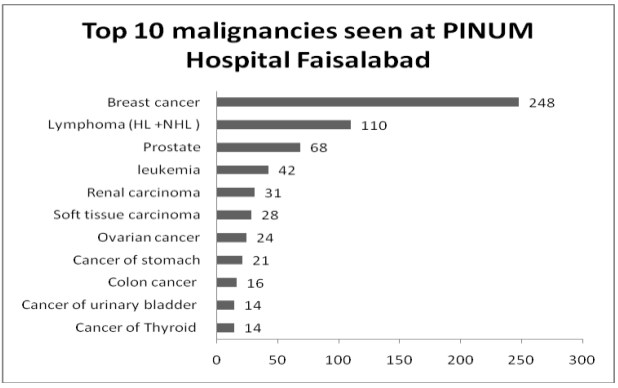


Figure.1: The top ten cancers at PINUM hospital

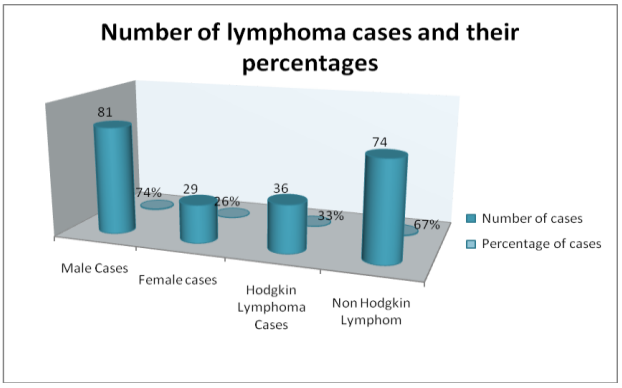


Figure :2 percentage and number of lymphoma cases by sex and types

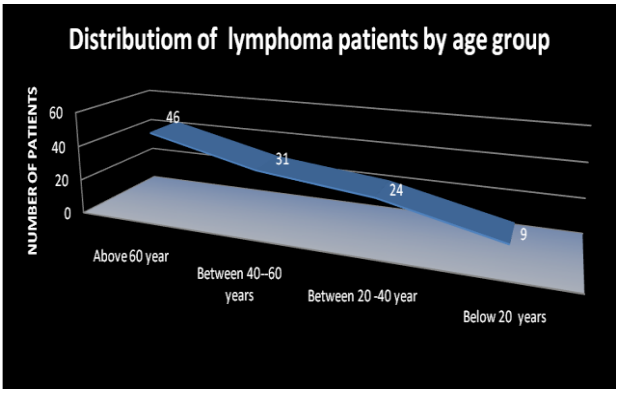


Figure :3 distribution of patient by age group

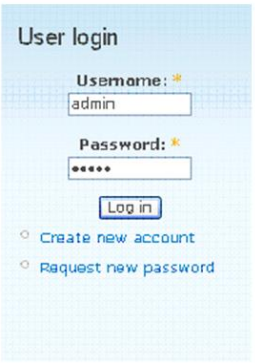
Development of the computerized hospital based lymphoma cancer database for PINUM Cancer Hospital Faisalabad was the first successful effort to develop a cancer database in Pakistan. This database will act like a one-stop-shop for the physician and scientists of the worldwide. In spite of the introduction of computing in medical sciences and emergence of bioinformatics in the country, the area of cancer database especially for the lymphoma cancer was untouched before the current studies developed lymphoma cancer database for PINUM Cancer hospital. The freely available lymphoma cancer database allows scientists and clinicians to search cancer-related data and information across the spectrum. All the clinical information can be stored and retrieved at the point of need. This database contained almost all the fields related to breast cancer risk as well as prognostic factors. This database will provide the management of lymphoma cancer data and replaces the previous paper based medical record with computer-based record system.

Features of lymphoma Cancer Database

- The Web-based data user friendly entry interface of lymphoma cancer database allows the physician and other concern person to enter the data independent of location.
- Data retrieval is much easier because of its simplicity and search ways that is by patient name, patient Identification number (ID) or by the date of registration.
- Access to the database is password restricted and thus is available only to allowed users.

Men are slightly more likely to develop DLBCL than women. Relative risk for developing lymphoma cancers with age factor was calculated. Lymphoma cancer risk was lower for persons with age below 20 years and persons who have weak immune system with age above 60 are at maximum risk of developing cancer. We have found that physically inactivity persons are at high risk for developing lymphoma. Pakistani population the majority of patients who developed lymphoma were totally inactive. Physically inactive persons are 4 times more at high risk of developing lymphoma as compare to physically active persons.

Fig.4:Punjab Institute of Nuclear Medicines(PINUM)



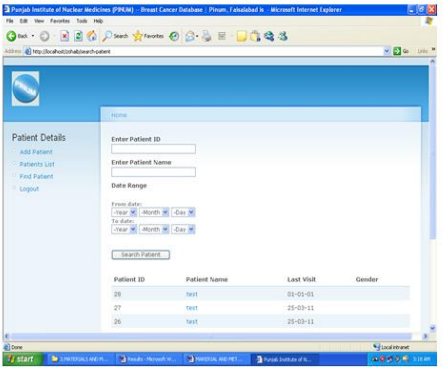
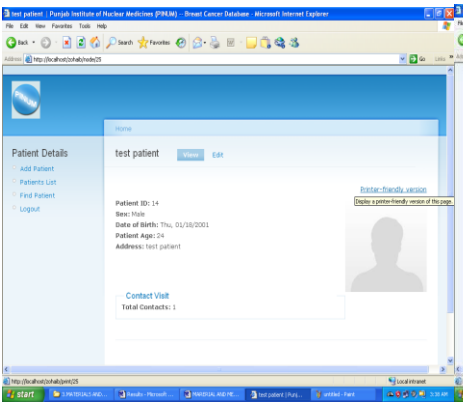


Fig.5: Password restriction **Fig.6: Data retrieval by three different ways.**

The effect of vigorous activities was also positive (i.e., reducing cancer risk); however, the estimates of RRs varied depending on the type of physical activity. Note, that while analyzing the effects of physical activity the bias could occur due to the difficulties in measuring this factor, its over reporting, and confounding factors. However, the inverse associations with physical activity (i.e., reducing cancer risk) have been described in other studies for most of human cancers, including colorectal, breast, and prostate.



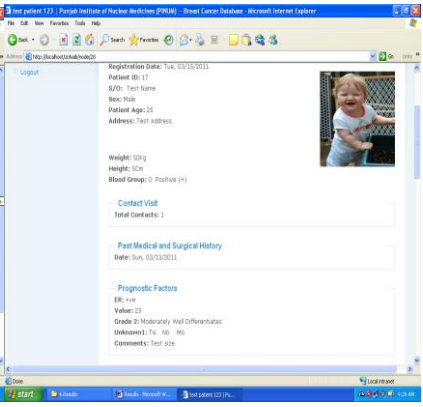


fig.7: printing option available in database **fig.8: View of registered data**

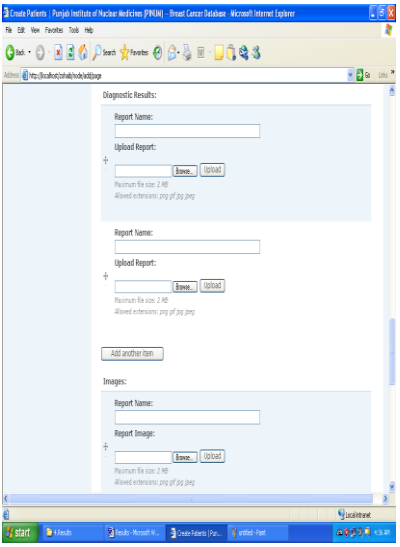


Fig.9: Facility of loading scanned tests and images

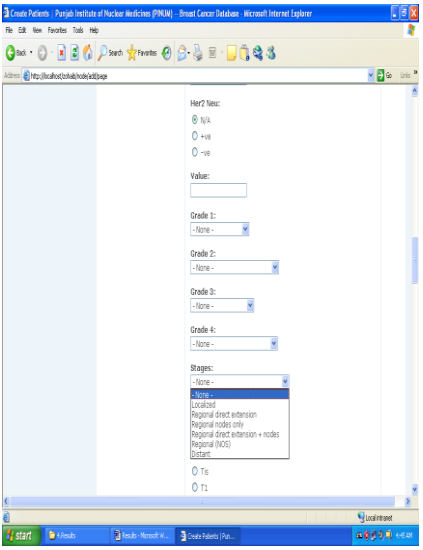


Fig.10: Easy input of data

Fig.11: user interface of lymphoma risk assessment tool for Pakistani population.

RISK FACTOR	ATTRIBUTES	PAR POPULATION ATTRIBUTABLE RISK	RELATIVE RISK	ODD RATIO
Socio economic status	Poor	63	1.34	1.9
	Middle	29	0.358	0.128
	Upper	18	0.195	0.038
Smoker /Naswar/Hoqqa	Yes	57	1.07	1.15
	No	53		
Drinker / alcoholic	Yes	41	0.59	0.35
	No	69		
Weight loss	Yes	43	0.64	0.411
	No	67		
hyper tensive	Yes	58	1.15	01.244
	No	52		
Diabetics	Yes	68	1.61	2.62
	No	42		
HCV + VE	Yes	31	0.39	0.153
	No	79		
Stomach infection / h.pyroli infection	Yes	21	0.235	0.052
	No	111		
Physically inactive	yes	88	4.0	16
	No	22		
Body Mass Index	Thin	51	0.86	1.072
	Normal	18	0.19	0.153
	Fat	31	0.392	0.41

Conclusion:

Lymphomas are a heterogeneous group of cancers that require the development of focused research and clinical approaches for specific histological subtypes. Integrating existing clinical, genomic and proteomic information provides a platform for examining biological variability in the pathogenesis of lymphomas and their responses to treatment response and will

promote the development of innovative treatment strategies.

Limitations and future research:

This data had limited value as the demographic details of the patients were partially recorded, lack of continuity and geographical variations within the country could not be clearly determined.

BIBLIOGRAPHY:

Aftab, K., Bhurgri, Y. and Pervez, S. 2006. "Small B cell Non-Hodgkins Lymphoma in Pakistan." *Pak Med Assoc,* 56(1): 22-25.

Beuhler, L.K. and Rashidi, H. H. 2005. *Bioinformatics basics; Applications in Biological Science and Medicine.* 2nd edition. New York: Taylor and Francis, 166.

Bhurgri, Y., Pervez, S. Bhurgri, A. Aaridi, N. Usman, A. Lag, K. Ahmed, R. Kayani, N. and Hasan, S.H. 2005. "Increasing Incidence of Non-Hodgkin's Lymphoma in Karachi, 1995-2002." *Asian Pacific Journal of Cancer Prevention* 6: 364-369.

Hodgkin. T. 1832. "On some morbid experiences of the absorbent glands and spleen." *Med Chir Trans* 17:69-97.

Huang, Taoying, Pareen J. Shenoy, Rajni Sinha, Michael Graiser, Kevin W. Bumpers and Christopher R. Flowers. 2009. "Development of the Lymphoma Enterprise Architecture Database: A caBIG(TM) Silver Level Compliant System Cancer Informatics." *Cancer Inform.* 8: 45–64.

Lucas, J.F., Bruijn, N.C., Schurink, K. and Hoepelman, A. 2000. "A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU." *Artificial Intelligence in Medicine* 19: 251–279.

Moss, Ralph W. 2004. "Galen on Cancer". Cancer Decisions. Moss in turn attributes this reason for the name to Paul of Aegina, 7th Century AD, quoted in Michael Shimkin, *Contrary to Nature*, Washington, D.C.: Superintendent of Document, DHEW Publication No. (NIH) 79-720, 35.

Oh, W.K., Hayes, J., Evan, C., Manola, J., George, D.J., Waldron, H., and Donovan. 2006. "Development of an integrated prostate cancer research information system." *M. Clin Genitourin Cancer* 5(1):61-6.

Reed, D. 1992. "On the pathological changes in Hodgkin's disease with special reference to its relation to tuberculosis." *John Hopkins Hosp Rep* 10:133-193.