

A Study on the Conceptual Frame Work of Dataware Housing in Health Sector in Pakistan

A Case Study of a Hospital System and Disease (Hepatitis C)

Anwar Ali Sathio*

Department of Computer Science, Faculty of CS & IT

Benazir Bhutto Shaheed University Layari, Karachi, Sindh, Pakistan

Email: Anwar.Sathio@bbsul.edu.pk, Anwar.sathiyo@gmail.com

Abstract

The decisions of people about medical treatments have a great impact on their lives. Health related sectors and patients often make decisions without complete understanding of the medical circumstances. The main cause is that health related data are available in fragmented, disparate and heterogeneous data silos. Without a centralized data warehouse structure to integrate these data silos, it is highly unlikely and impractical for the users (of concerned departments) to get all the information required on time to make a correct decision. In this research study, a case study has been discussed hospital system and a disease (Hepatitis C only) through literature review, it was found that in government hospital the data recorded analyzed in old fashioned, manually where as in private hospitals there is computerized Database to record the patients data, but failure is that there is no coordination between government and private hospitals to share the patients history and also no concept of Database overall in government hospitals in Pakistan. After all, in this research study, a conceptual frame work has been discussed and proposed model for implementation of Dataware house in health sector to centralized the data for analysis and decision making. Outcomes of this research are: one is; documenting the comparisons of data warehousing architectures, logical and conceptual data warehousing models, other is proposing a data-warehousing model (taking one disease Hepatitis C) as model for implementation which is suitable for Government Health Sector as well as other concerned departments in Pakistan. ext fonts are prescribed; please do not alter them.

Keywords: Clinical Data Integration; Clinical Data Warehousing; eHealth; Electronic Health Record.

* Corresponding author.

1. Introduction

This research study is for to design a conceptual frame work for implementation of e-Health data ware house in government sector. Actually, this study focused on the conceptual frame of data ware housing to store the record of patients of all groups, levels who visits the hospitals, dispensaries, BHU etc daily. This is very necessary to have the update record of the patients at all levels to make decisions such as for planning and developments of health projects, funds allocations, daily , weekly, monthly ,quarterly records of patient related information. So, reason behind the study to convert the old system with new innovative and integrated metadata for decision making and all other required needs for their developments. In Pakistan, the record keeping system of patients in government owned hospital are still going on manually. All the analytical information analyzed through old fashion. The ration of budget for health sector derived through manual system. All the data derived through a procedure, from Basic Health Units, Taluka/Town/Tehsil hospital, district hospitals special hospitals and big city hospitals. All research work done through these hospitals. Budget allocation system for health after getting the information and estimation of health department. As the manual system of recording the information of patients are very slow and inaccurate, inconsistence that's why the correct information of patients are so difficult. Another reason behind this is having different levels of medications in the country such as allopathic, homeopathic, self old medication, unauthorized and un recognized medication and other ways of medication prevailed in the country but having no way of central record of concerned persons. There is no any platform to keep the record of all patients ,doctors and legal medical staff. This is very necessary to have the full electronic record of medical concerned even pharmaceuticals. In Sindh, Karachi mega city was studied to get the information of patients record keeping manual in the hospitals. It was shocking that still today the record keeping system in the government hospitals are running manually, there is a record room which has responsibility of daily record, weekly, monthly record keeping system. Then summarized data prepared and dispatched to up levels of health departments for estimations and budget allocations. This kind of approach is like ill approach. As all private hospitals in the city Karachi such as Agha Khan Hospital, Ziauddin Hospital have been equipped with latest information technologies and running with RDBMS to keep the record of patients and resolve the issues at home level. But there is still need to enhanced that system and merge the two systems; government and private hospitals with centralized information systems for getting the timely information of the areas. It will help the government to focus the red alert issues and also can make good decisions for the betterment whole population of Sindh. This study encompasses and suggests the conceptual framework for implementation of Data ware house in health sector in Sindh, Karachi city was selected for study. The benefit of health Data ware house is multidimensional and heterogeneous; these will effects good impacts over the development in health sector and also minimize the cost of health expenditure of common man, and the benefits of this will fruitful for all.

1. Proposed conceptual framework for health Dataware Housing for decision making for betterment of government Health Sector with complete information of patients for planning and analysis.
2. This will minimize the risk in Health sector to combat in mean time in case of any viral infection in the country.

3. The record of each patient who visiting public and private hospitals to be saved at one place for future planning and health improvements in the country and can be utilized for analytical and statistical use.
4. The basic aim of this research study is to design a conceptual frame work of data warehouse in health sector for delivering the information to patients, civil as well as government organization, locally and internationally to communicate with them through delivering the analyzed reports in case of disaster, viral attacks and other necessary purposes.

2. Literature Review

This section presents how IT is used in modern health care service provider organizations and the main problems related to the adoption of IT. The recent developments in health care have heightened the need for application of IT. The field of science which studies about the application of IT in health care is known as Medical Informatics or Health Informatics or E-health. Recently, many healthcare organizations/hospitals have been interested keenly to adopt IT into their services process.

According to Goldstein medical informatics defined as “the interconnection of medical, business, centrally consumer for improvements in both quality and cost effectiveness in healthcare effectively”. The medical tools are formal medical languages, formal standards, and operability and information system for healthcare. The information regarding medical records of hospital are varied; including personal data of patients, letters from different doctors; diagnosis, laboratories reports, nursing observation notes and administrative information [1] One question that needs to be asked is about the approach to convert massive piles of paper based data into electronic format.

According to authors [2] the ideal Institute of Medicine has suggested six basic objectives of improvements in healthcare through IT utilization, these six aims are:

1. Safe: keeping safe the patient from injuries.
2. Effective: through proved tools of knowledge, with quality service provision.
3. Patient-centered: assurance of services to all with preferences.
4. Timely: Service to all patients without delay.
5. Efficient: keeping high standard with best services, supplies, equipments, energies and ideas.
6. Equitable: equal opportunity and service provision without any carelessness to everybody.

As mentioned in above six aims, the proper implementation and use of IT in health care organizations can drastically improve the efficiency of medical decision making by providing relevant data. In order to make patient data ubiquitously available worldwide for health care as well as for clinical research, epidemiological research, bio-surveillance and population health reporting, E-health ignores institutional boundaries and tries to

share patient data among institutions according by authors [3]. The key problem with this explanation is that it has not considered the fact that still the health care institutions are not well connected to share the data. According different HER researchers, the main challenges in medical informatics are:

- Massive implementation and maintenance costs of health information systems
- Incompatible health information technology systems and proprietary software
- Lack of clinical data standards for sharing information
- Application systems which are not integrated to work collaboratively
- Medical Information Fragmentation at latest heterogeneous and disparate information systems[4].

Since the research presented in this thesis is about integrating fragmented electronic health records through a centralized Health Data ware House only the last mentioned challenge is described in detail in this literature review.

In order to facilitate the process of information, multiple information systems are used by healthcare organization for getting best decisions. The important thing is that to keep maintain the consistency of medical information, however information system developed in different technologies by different venders [3], this make difficulties for information flow. The many venders providing solutions to minimize the difficulties between different environments of information but still healthcare system facing many problems. Such as decentralization of information access, various integration interfaces from system to system and increases of system interfaces as due to geographical increases.

The distributed and heterogeneous systems such as HIS, RIS, PACS; they are working across different locations of the hospitals, data with these systems controlled by different medical staffs dealing the privacy of data and confidentiality, keeping safe from others to use. According to authors [5], the medical data is a island of data residing at various places , access of that data is time consuming job, but through proper technologies it can be achieved. The books and recent research papers reviewed provide concrete evidence that the fragmentation of medical data across multiple disparate and heterogeneous information systems is a real problem yet to be solved completely. According to shepherd un availability to synthesis information from disparate information sources both the patients and practitioners of healthcare will not be able to get benefitted through stored data, health information is complex data must be available at different levels. Medical data integration attempts fail due to lack of interest of all stacked holders of healthcare.

2.1 Electronic health records (ehrs)

One of the most significant current discussions in health sector is EHRs. Proper documentation of the health status of the patients is very important. It is becoming increasingly difficult to ignore the importance of having centrally managed and shareable patient health data. Traditional paper based method which has been used for centuries has many drawbacks. When implemented properly, EHRs can support the current information requirements of the health care industry. However, far too little attention has been paid to EHRs by both patients and health care organizations. This section presents an overview about what an EHR is, key benefits to health

care organizations and patients and the problems associated with the use of EHRs. According to [6] an EHR can be defined as: “For clinicians, Electronic Health Record (EHR) is a way to secure, virtual on line, point-of-care, patient centric information resource, aids decision of clinicians regarding patients when and where required tally workflow and provide quick response for to patients and supports to collection of data pertaining to resource planning, disease surveillance and reports.

The main causes relates with HER according to authors[2,7] tangible and intangible benefits associated with EHRs to health care providers and patients are:

1. Quality care improvements
2. Minimizing the errors through changing the old manual system
3. Patients’ information intake all times.
4. Save the history of patient’s medication.
5. Provide the satisfactory required reports for the organization.
6. Patients information easily accessible/shareable when required by concerned.
7. Ease the Admin by minimum efforts.
8. Decrease the expenses of the organization.
9. Make the entire satisfaction for hospitals/ health department.
10. Reduce the cost for health care.
11. Improve the progress of health workers.
12. For satisfaction of patients, doctors, organization as well as government.

The manual records systems can be used only one person at a time, it make difficulties to quick access, store, and retrieve of information. The electronic records make it easier to access, long time secure, easy to store and retrieve the required information within few seconds. EHR is now very popular in healthcare fields, to provide quicker access information to all corners of concerned of healthcares. According to authors [8] data is used and organized in EHRs according to operational purposes for overview of health status of patients etc.

Although EHR systems are increasingly used in many health care organizations, there still remain some barriers to their adoption.’ There is a huge gap between the availability of innovative technologies and their application in daily health care. Reasons for this include the inherent complexity of the field, costs, as well as ethical and legal requirements.’ [3]. The main barriers to the implementation of EHRs as identified by [3,4] are:

1. Cost of Conversion

The cost involved in converting the paper based records systems to an electronic system is massive. This cost includes the acquisition of hardware, software and trained technicians. Certainly the biggest impediment is financial since most of the small scale healthcare organizations are unable to bear the above mentioned costs without the help of a funding agency.

2. Concerns about Privacy and Confidentiality of Data

The provision of services by much health care organization including all facilities, in the distributed environment, EHRs are always on the move, and the security of critical information infrastructure becomes more difficult. Main implication of this is the need to follow stringent standards. Special cultural environment in a healthcare organization demands security to be a fine-tuned balance between technologies, human elements, standard practices and procedures [9]. In many instances the same data set of a patient is accessed by administrative staff, physicians, nurses and laboratory in order to make decisions regarding patient's healthcare. Therefore all those who are involved in the process have shared responsibility and accountability to maintain the security and integrity of a patient record.

3. System Interoperability/ Intercity Hospitals networking for information sharing

Various clinical and administrative systems within and beyond a healthcare organization must work together in a smooth manner to give optimum performance. But this does not happen since most of the proprietary software systems by various vendors do not communicate with each other effectively. The health relating data mostly in all formats, whether on paper or electronic format, are trapped in multiple data silos in multiple vendor products. If patient recorded at advance EHR, it can be accessible at other places where it required or needed [4].

4. Shortage of a Well-Trained Medical Informatics Workforce to Lead the Process/ Untrained staff for data staging and general medical report requirements

To maintain an efficient electronic health record keeping system, an organization must have a well trained workforce with a clear understanding of the requirements of both the worlds of medicine and IT. They should be highly motivated to carry out operations and make innovations to support rapidly changing requirements of healthcare industry. This type of a workforce building is a highly time consuming and costly effort.

5. Data Storage Requirements

The need for data storage in healthcare grows rapidly. Systems such as Picture Archival and Communications Systems (PACS), which handle digital X-ray, CT, and MRI images use significant storage. These systems are frequently used by most of the healthcare organizations today. Improving the storage environment means more than simply adding better storage hardware. Centralized, standardized storage management software, which are independent of hardware and are able to manage the diverse, heterogeneous environments that exist in real-world data centers, are important ingredients in the ideal storage prescription. This also should include

maximizing utilization of existing storage, improving backup and recovery performance, and classifying structured and unstructured data to improve archiving

and retrieval. This last benefit is of particular importance to clinicians, who need the right information at the right time. [9].

6. Fragmentation of EHRs

EHRs (Electronic Health Records) are fragmented across various disparate and heterogeneous organizational systems. They are fragmented among and within hospitals. Health care practitioners, providers and patients often make decisions about medical treatments without complete understanding of the circumstances. The main reason for this is that medical data are available in fragmented, disparate and heterogeneous data silos. Without a centralized data warehouse structure to integrate these data silos, it is highly unlikely and impractical for the users to get all the information required on time to make a correct decision. The active participation of leadership, clinicians, patients and all those involved in healthcare is a must to make EHRs and information technology an essential resource to solve the problems in medical field. This suggests the need for every stakeholder in health care to participating in implementing working framework for EHRs. Benefits of adopting EHRs in clinical practices largely outweigh the efforts required to overcome the barriers. According to Bill Inmon [10] “ A data warehouse is a data repository which inclusive with data as domain oriented, integrated, permanent and time-variant set of data for helps the managers to good decisions.” .

- a. **Data Integration:** As the data come into data ware house, its image would be physical, as these data sources may have separate and several channels passing through legal way of formation, conversion and summarization [10].
- b. **Permanent Data:** When updated the record a snapshot rewrite over previous data, its history saved, its only when changes occurs [10].
- c. **Time variant:** The accuracy of data in each moment is compulsory, a form of time marking to keep the record of Time about accurately recoding the data [10].

The management required the data warehouse which has to provide functions as :

- A. A shadow of rules which useful for making decision in big organization.
- B. Data collection point which can manage the process for data acquisition.
- C. It is a strategic data store.
- D. A source of stable data.
- E. Data ware house should support ad hoc queries additionally.

The important fact about data warehouse is that having the solutions of information overload. There are following some areas considered as important by author [10] which are:

- subject defined
- Multiple source integration

- Data Analysis in time
- Support the Adds ad hoc queries and reporting
- Capabilities of analysis
- Decrease the load on technology of information
- Supportive in all complex queries
- Decrease the load of transaction of databases
- Supports planning process continuously
- Supportive in corporate data conversion into strategic information

The Data ware house has few drawbacks as its very expensive, high failure rate along with other many qualities such as high needs, powerful sponsors, and little time scale. Following are the important factors regarding the cost of data warehouse:

- Spent time in careful analysis of needs
- DWH Design and implementation efforts
- Equipments, Hardware expenditure
- Application Software expenses
- Charges of maintenance and support
- Expenditure regarding Resulting re-engineering efforts

There are three basic approaches in data ware house model such as relational, multidimensional modeling or hybrid. The each model approach has three basic levels by authors [10] which are as below: [11]

A *Levels In Dataware House:*

- **Semantic level (Conceptual):**

An abstract level which represents matters in investigation, thoughts, concepts and domain analysis can be done technologies of software in thinking.

- **Logical level(Applications):**

It's a model by logics to show entity relationships to encompass data integration, just real existing objects will be considered.

- **Storage level (Implementation):**

It's a actual representation of tables in storage level within existing database through physical data model for all users who need data for use. [11]

2.3 Why A Need For A Health Data Warehousing In Health Sector

Electronic Health Records (EHRs) which keep the records of patients are normally retained in the hospitals, but it may be possible for patients to be treated in different hospitals required health records from distinct places of patient medical history. It's better to have right information, in right time through secure, safe and reliable sources for information exchange by machines.

This data must be in distributed and also heterogeneous databases with many varieties data types etc available in medical subjects Masseroli and his colleagues, there is a risk of failure of data ware house due to not meeting the required need of enterprises [12]. Health care organizations require data warehousing solutions in order to integrate the valuable patient and administrative data fragmented across multiple information systems within the organization. As stated by Kerkri and his colleagues a technical level, information sources are heterogeneous, autonomous, and have an independent life cycle. Therefore, cooperation between these systems needs specific solutions. These solutions must ensure the confidentiality of patient information. It is a repository of health related data, by this medical care organizations get access to patient care data, and enhance quality of decision making for users. Hospitals needed effective quality treatment including reducing costs [5]. Medical domain has certain unique data requirements such as high volumes of unstructured data (e.g. digital image files, voice clips, radiology information, etc.) and data confidentiality. Data warehousing models should accommodate these unique needs. According to [8] the task of integrating data from several EHR systems is a hard one. This creates the need for a common standard for EHR data.

2.4 Case-1 Study of Government sector record keeping manuals

The paper under discussion was a research study conducted to check the government hospitals of Karachi by authors. [14] It was conducted few government hospitals such as Jinnah hospital, national child hospital and lyari hospitals. Study covered many aspects through questionnaires as mentioned in table No.03 [14].

Table 1

	CHK	%	JPMC	%	LGH	%	NICH	%	Total	%
Not Present /Shifted elsewhere	1	0.7	19	14.0	1	0.7	0	0.0	21	15.4
Available but not readable	0	0.0	1	0.7	0	0.0	0	0.0	1	0.7
Available and readable	54	39.8	26	19.8	23	16.9	11	8.1	114	83.9
Total	55	40.5	46	33.8	24	17.6	11	8.1	136	100.0

In table No.01 , it is clear that status of record storage , where it is accessed that only eighty percent record can be in safe position where remaining data was not access able.

Table 2

TOTAL	CHK	%	JPMC	%	LGT	%	NICH	%		
Information absent/ Not available	1	0.7	19	14.0	1	0.7	0	0.0	21	15.4
Upto 25% Information Available	19	14.0	12	8.8	13	9.5	2	1.5	46	33.8
>25%----75% information Available	6	4.5	5	3.6	4	3.0	1	0.7	16	11.8
>75%----99% information Available	24	17.6	8	5.9	6	4.4	7	5.2	45	33.1
100% information Available	5	3.7	2	1.5	0	0.0	1	0.7	8	5.9
Total	55	40.5	46	33.8	24	17.6	11	8.1	136	100.0

In table No.02 the detail char explained that what kind of information displayed in stored record ,it stated that very few areas were covered in records.

The main findings of the study conducted by authors [14] were following:

- No proper record keeping system was available.
- No any proper post established for record maintenance
- 90 percent hospitals' departments/units did not have the computer record.
- Availability of record not more than a few years
- The retrieve of old record could not be accessed in minutes
- Proper OPD records not easily available /improper /incomplete record in all units.
- No proper place available for record of patients.
- As per study results the 56 percent record kept in either store rooms or in nursing duty room
- Height and blood pressure record was missing
- No any trained staff available for keeping the patients full record daily properly.
- Doctors also did not pay attention towards the record of the patients.
- Supportive and extra related Information of patients missing.

The books, research papers, white papers, government reports and websites referred in the preparation of this literature review provide a clear idea about the field of medical informatics, unsolved challenges in the domain and the technological and other concepts which can be helpful in dealing with the challenges. This research is aimed at proposing and developing a solution for a main technological challenge. The challenges is that how the efficient implementation of health information technology will make the Government Health sector into moderate system for planning and development of the Healthy society; thus, save the time and budget well. Though the field of medical informatics emerged about four decades back, it had proved how useful it can be for providing better health care. One of the main building blocks of medical informatics is the Electronic Health Record (EHR) which is used for recording patient information electronically. There is a demand to get in timely information for analysis of health system and patient's records etc need a well organized Database System with latest technology to keep the archived record for future use in planning and development of health sector. There is world many experiences available such as potential of EHRs have not been realized yet due to various technical and nontechnical issues in government sector. The One major obstacle is Government Lack of initiative to the fragmentation of health data across multiple disparate and heterogeneous information systems. This is the research problem discussed, research papers reviewed to conforms that health Dataware housing is still an unresolved problem in health sector. This problem requires solutions technically as well as nontechnical. The main focus research is initially narrowed into developing a designing the conceptual frame work model for health sector, selecting the two case studies in Karachi city. Without integrating the information within individual organizations, national level integration cannot be achieved. This research will investigate about the application of data warehousing and integration technologies into medical domain firstly designing the conceptual data model for Data ware housing in Health sector. Internationally, SAS® data integration technologies and open EHR architecture can be used to implement a solution for the research problem. SAS® is accepted all over the world as the leading business analytic tool and open source data integration tools are gaining popularity. The research problem focused on is a challenge due to various reasons. Some of them are, heterogeneity of medical information systems, lack of widely accepted standards, incompatibility among proprietary software and lack of support from patients, clinicians and policy makers.

3. Research Design And Methodology

3.1 Characteristics Of The Research Undertaken

1. Propose a data warehousing model to integrate clinical data derived from disparate systems into Hepatitis C(HCV) data.
2. In keeping with Applied Science processes, in order to solve the problem, well known and accepted theories, principles and tools relevant to data integration, data warehousing, medical informatics and medical data interoperability are employed.
3. The outcome of this research has immediate applicability in the problem domain.
4. There is no need to make changes in proposed solution.

The main aim of this research is to propose an applicable model solution for a practical problem which is suitable for immediate use.

3.2 System analysis layout for conceptual design of Data ware Housing

In this section we have the examples in the world of research about health informatics and data ware house , this search will help to design the model of health data ware house, to investigate the Current State of Medical Data Fragmentation, Relevant Data Warehousing and Integration Technologies

This stage's outcome consists of:

1. Studying the structure of data distribution and communication in existing health information systems.
This facilitates the understanding of the underlying data related problems in the health care information systems.
2. Compare and analyze various data warehousing technologies in order to identify the most suitable model(s) to health care domain.
3. Compare and analyze various data integration technologies.

The findings of the processes in this stage are important to understand the data warehousing and integration technologies employed in various other fields and their relevance to health care domain, through search of published articles, papers etc. The Current State of Health Data, Comparison and Analysis Results of Data Warehousing and Integration Technologies. In past we can examine the data tables of cardiac surgery data using the data manual published by Australasian Society of Cardiac and Thoracic Surgeons (ASCTS) under ASCTS National Cardiac Surgery.

3.3 Research Methodology

1. Proposed conceptual frame work and Model design of Health Dataware Housing in Health sector

There are several levels for design of Dataware housing technology, such as development cycle, software selection, hardware configuration and networks , as centralized or distributed database systems.

4. Discussion and results on proposed conceptual Dataware house model

The following is the description of a sample model, before understanding of the model ,the HCV must be explored , in Sindh context, so that a model may be finalized.

In this regard, sample was taken from cardio surgery model, the details of tables which can be analyzed through the understanding of tables and their meanings with the data type applied on it.

4.1 Hepatitis C virus overview

Hepatitis C is a viral infection of the liver which had been referred to as parenterally¹ transmitted "non A, non B hepatitis" until identification of the causative agent in 1989. HCV is a major cause for hepatitis and liver's chronic disease such as liver cancer , and cirrhosis [15].

A. Pathogen view:

HEPATITIS C VIRUS (HCV) is one of the viruses amongst A,B,C, D or E, which enveloped RNA virus in the flaviviridae family, humans and chimpanzees are only known for infection with similar disease. In its genome high propensity of inducing chronic infection, resulting a cluster of HCV into many genotypes which determine the severity of disease [15].

B. Infection features:

The HCV infection observed within 15 to 150 days period, most common symptoms are fatigue and jaundice.

C. Ways of HCV transmission into masses:

The following are means which can cause the transmission of virus:

1. Direct connection with human blood
2. Reuse of sterilized needles, syringes, and medical equipments etc.

D. How HCV can be diagnosed?

The lab test are the only way to identify infection more than 95 percent, which are:

1. Enzyme immunosorbant assays (EIA)
2. A recombinant immunoblot assay (RIBA)
3. RNA (e.g. polymerase chain reaction or PCR, branched DNA assay) [15]

4.2.1 An overview of Sindh context regarding HCV

After review of research studies it is found that Pakistan is a country of 170 million masses , where people facing Viral Hepatitis disease rising as a significant health problem[16]. In current situation , its estimated that the prevalence is tripled after every decade [17] The aggregated data regarding the HBsAg and Anti-HCV prevalence and its mean results after study of 203 small research studies at Table No.03 [17].

4.2.2 Diagnosis for Hepatitis B, C in Labs

Hepatitis can be diagnosed through variety of laboratory tests, in early test, which to be performed is an ELISA test for antibodies or viral antigens to identify in the blood whether it's present or not.

The very common viral tests for Hepatitis virus which are:

- (A) Hepatitis B: HBsAg.
- (B) Hepatitis C: Anti-HCV

If the results are borderline or positive, or if the doctor has a strong suspicion for infection, a PCR test may be performed, which detects even very small amounts of viral particles in the blood of the patient.

Table 3

S.NO	PATIENTS	TEST (HBsAg)	TEST (Anti HCV)
		PERCENTAGE (%)	PERCENTAGE (%)
1	Children	2.3	2.5
2	Pregnant women	2.5	5.2
3	General population	2.6	5.3
4	Army Recruits	3.5	3.1
5	Blood Donors	2.4	3.6
6	Healthcare worker	6.0	5.4
7	High Risk Group	13.0	10.3
8	Patient with provisional Diagnosis	12.3	12.0
9	Patients with chronic liver disease	25.7	54

4.3 The Data Tables

List of table names taken into consideration for databases to record the patient’s whole data, which can be useful for Health Dataware housing they are: [19]

1. Patient Demographics
2. Patient Risk Factors
3. Pre-Operative Status
4. Previous Interference
5. Haemodynamic Data
6. Operative Status Category
7. Minimally Invasive
8. CPB and Support

- 9. Liver Transplantation
- 10. Liver Surgery
- 11. Post-Operative Data
- 12. Post-Operative Data - Complications
- 13. Mortality/Readmission
- 14. Automatic Data [19]

Table 4: Showing the Age, Gender Distribution and Etiological Factors for HCC in Pakistan, 1970-2011.[18]

SUIT AS et al (27_1999-2011)	Study period
645	Sample size
Karachi	Location
546	Male No.%
99	Female No.%
56.93±11.15 (18-95)	Age Meant (range)
119	HBsAg+ (%/)
371	Anti HCV No(%)
26	HBsAg+ Anti HCV+ No(%)
99	HBsAg-ve AntiHCV-ve
15	HBV+ HDV+No (%)
15	B+C+D

4.4 The Register Table Structure

The demographic structure of the patients record displayed in the table No.05.

4.5 Proposed data models and data warehousing model

The following are model which can be suitable design for data warehouse models:

4.5.1 Data Model for Patients

It was found that there was no any DBMS available in the government sector, all the data recording of patients doctors and medicine on manual registers. The following Data Model has been initiated for the government held clinics, BHU, Dispensaries, and Hospitals. The detail of tables and diagrams are as under:

Table 5: Patient Demographics [9]

Column Name	Meaning	Data Type
PCNIC	Patient CNIC or his /her parent/ relative	integer
RELpatient	Relation with patient	varchar
LNAME	Patient last name	varchar
FNAME	Patient first name	Varchar
MNAME	Patient mid name	Varchar
DOB	Date of birth	Alphanumeric
Age	Of patient	Integer
SEX	Sex of patient	Varchar
MStatus	Maternal status of patient male/female	Varchar
MEDC	Medical ID number of patient	Alphanumeric
ADDRESS	Patient address	Alphanumeric
Village/ mohalla/street/ward	-	Alphanumeric
UC	-	Alphanumeric
Tehsil/Taluka/subdivision	-	Alphanumeric
District	-	Alphanumeric
Division	-	Alphanumeric
Province	-	Varchar
Country	Patient native, foreigner, etc	Varchar
Post Code	Postal code	Alphanumeric
PHONE1	Tele number	Integer
PHONE2	Tele number	Integer
CELL 1	Mobile number	Integer
Email	Email of patients caretaker relatives	Alphanumeric
RACE1	Race of the patient	Varchar
CASTE	Caste of the patient	Varchar
PMRECforSUR	Patient medical record of surgery	Varchar
ADforSUR	Admission date for surgery	Varchar
DDofPatient	Discharge date of patient	Varchar
PINSUR	Patient insurance if any	Varchar
PPHist	Patient payments history record /status self, govt.s support, welfare org, other records	varchar
PDOP	Date of first diagnosis /operation take place	Alphanumeric
DOCstaff	Doctor / staff look after the patient	Varchar
PresCrip	Medicine record	Alphanumeric
TestREC	Patient chemical test record, Xrays, etc	Alphanumeric

4.5.2 Data Warehouse model in health sector

It is proposed that the patients record keeping system in the government hospitals has been taken and derived from the Records of Cardiac Surgery Clinical Data stores. A data warehouse can be used to integrate the fragmented data sets, after creation of Datawarehouse, it will be maintained with data through processes which derived from the data by Extract, Transform and Load [3,10].

4.5.3 The Concept of Meta Data

It's a data about the data in which user can navigate, it's an effective use of Data warehouse, it's just like an index of contents in data warehouse. It's like tracking watch over the data warehouse.

4.5.4 Proposed Conceptual Data Warehouse Model Design

The suggested conceptual data model design is to be Object Oriented Multi Dimensional Modeling, there is several reasons to select the multidimensional modeling, in which facts are represented by Fact tables and Dimensional tables, it also uses a generalized and specialized relationship to categorized entities that contain subtypes. Making Health data warehouse, this research proposes a conceptual framework and data warehousing model for integrating disparate and heterogeneous hospitals data regarding patients diagnoses and medical treatment, to integrate fragmented data future use. By this, improvements occurred into information quality, this data source may be internal or external, in different formats, structures. This data is consolidated and integrated into single repository

When a Doctor, Nurse, Patient and administrative staff of the hospital tries to obtain a report from a fragmented data set they have to face the three challenges listed below:

1. Need to obtain the service of programmers/analysts to locate, analyze and compile the data.
2. Problems arising due to data inconsistencies.
3. Navigational access to all data atom with definitions.

These processes are carried out properly, but the report will end up creating yet another level of confusion and will lead to misinformed medical decision making. Even though the report generation program should be simple to write, retrieving the data for the report.

4.6 Proposed Star schema for Data warehouse model in government Health sector Hospitals as model

The Health Data ware Housing Model developed in this research is a highly customizable model. Star Schema can be customized to accommodate various data retrieval needs. Health Data Warehousing is a cost effective data integration methodology. In order to make use of the model presented in this research in a clinical environment, the conditions mentioned below should be satisfied.

There should be a document which clearly defines the structure of the data tables currently used by the Government hospitals in the city Karachi. It should clearly know what the data retrieval operations are going to be executed when using the data warehouse.

A. Experiment:

The method for data processing proposal, getting UCI repository dataset of patients records [<http://mldata.org/repository>], which donated by Josef Stefan Institute in Ljubljana. *Hepatitis* (Greek) substituting the meaning as 'liver' and suffix –it is denotes 'inflammation' of the liver. There are five kinds of the virus, as hepatitis A (HAV), B (HBV), C (HCV) D (HDV) and E (HEV).

In example dataset is used to predict patient's chance of survival. The dataset contains 155 records, of which 32 patients die and 123 survive. There are 19 attributes – 6 nominal and 13 numerical. [19] All symptoms in patient's record are:

B. Data Set, Attributes information:

1. Class: DIE, LIVE
2. AGE: 10, 20, 30, 40, 50, 60, 70, 80
3. SEX: male, female
4. STEROID: no, yes
5. ANTIVIRALS: no, yes
6. FATIGUE: no, yes
7. MALAISE: no, yes
8. ANOREXIA: no, yes
9. LIVER BIG: no, yes
10. LIVER FIRM: no, yes
11. SPLEEN PALPABLE: no, yes
12. SPIDERS: no, yes
13. ASCITES: no, yes
14. VARICES: no, yes
15. BILIRUBIN: 0.39, 0.80, 1.20, 2.00, 3.00, 4.00
16. ALK PHOSPHATE: 33,80,120,160, 200, 250
17. SGOT: 13, 100, 200, 300, 400, 500,
18. ALBUMIN: 2.1, 3.0, 3.8, 4.5, 5.0, 6.0
19. PROTINE: 10, 20, 30, 40, 50, 60, 70, 80, 90
20. HISTOLOGY: no, yes [19]

4.7 Proposed Data Model Of A Hospital

The following model is a proposed model for a hospital, clinics and even basic health units in the country at Fig. No. 01. It is very precise and smart model which can be applied in any case.

4.8 Propose Health Data Ware House Model

It is a concept which may be helpful to record the patients all data under one place with sharing of NADRA to build a Health data warehouse of the country. It will help to take decision regarding the development of health standards in the country and budget allocation, disease analysis and viral outburst due to disaster when occurred in the country. It will also helpful to keep the record for identification of person when a body could not be identified and also helpful for crime control system in big cities. The proposed Health Data warehouse mentioned in the Fig. No.02.

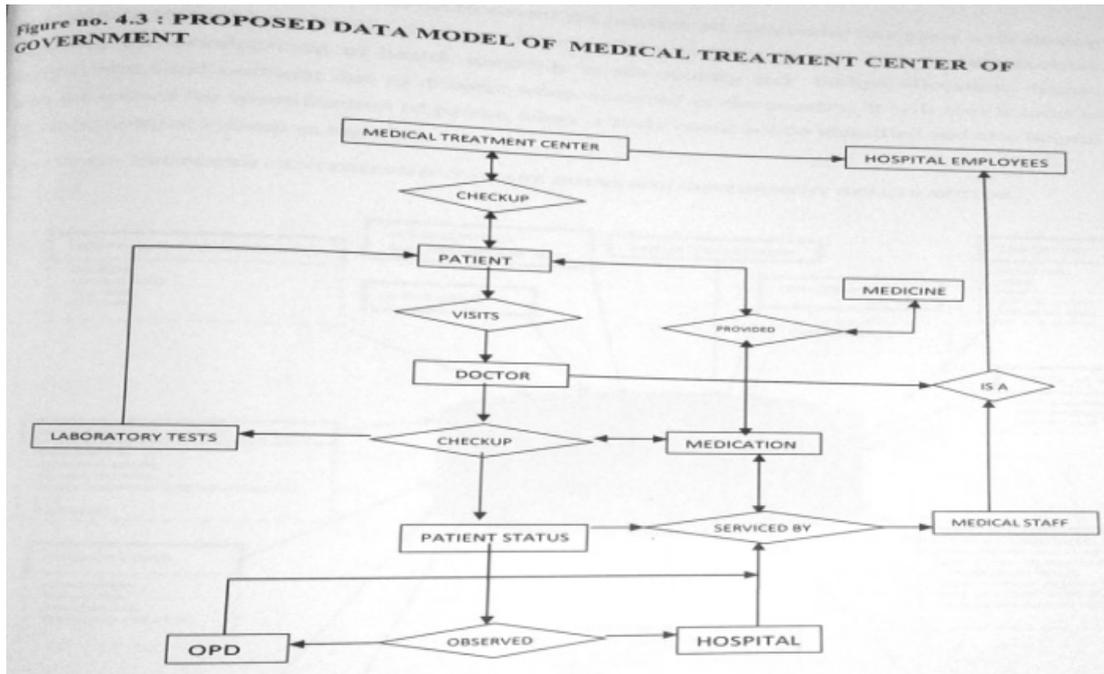


Figure 1: proposed model for a hospital, clinics and even basic health units in the country

4.9 System of Data Collection at All Levels and Channels

- a. The following chart showed the position of reporting systems for data tabulation and analysis at different levels, it was observed during the survey and interview of 20 hospitals of city Karachi and Tando Muhammad Khan.
- b. The following chart states the performance comparison of Dataware house with other offline databases and record keeping manuals in Sindh. The ratio of accuracy and human error were assumed to show the comparison with Dataware House in Health sector.

4.10 Limitation of proposed study design and impacts

It is very difficult to complete error free design whether statistical design ,or modular design of studies. No one can make an error free and hundred percent complete design in the science studies. In this conceptual model design for Health Data Warehouse, the most important factor is the structure of Health data in the form of HER

and its availability round the clock analysis. The next step is set proper and accurate patient's data availability or inputs for Health data warehouse. Unfortunately, data integrity of patients in all sectors of hospitals in the country like Pakistan is very big task to come to be true in future. The resources for data collection from disparate sources such as OLTP, other ONLINE data communication between the source and destination places, as from gross root level to big organized hospitals, including private hospitals, it's a big problem in Pakistan. Some hospital deliberately hides the facts and figures to be recorded in time due patient's private policies. There is lack of resources such as necessary devices, communication systems, digital laboratories, online availability of medical test systems and technical trained staff, these are the main factors involved to resist the accurate and patient data in time. In that perspective, it is mentioned here that before implementation of this conceptual design since for just on disease HCV taking as model, and all other spreading viral disease be streamed and synchronized into design as Data marts as sources permit.

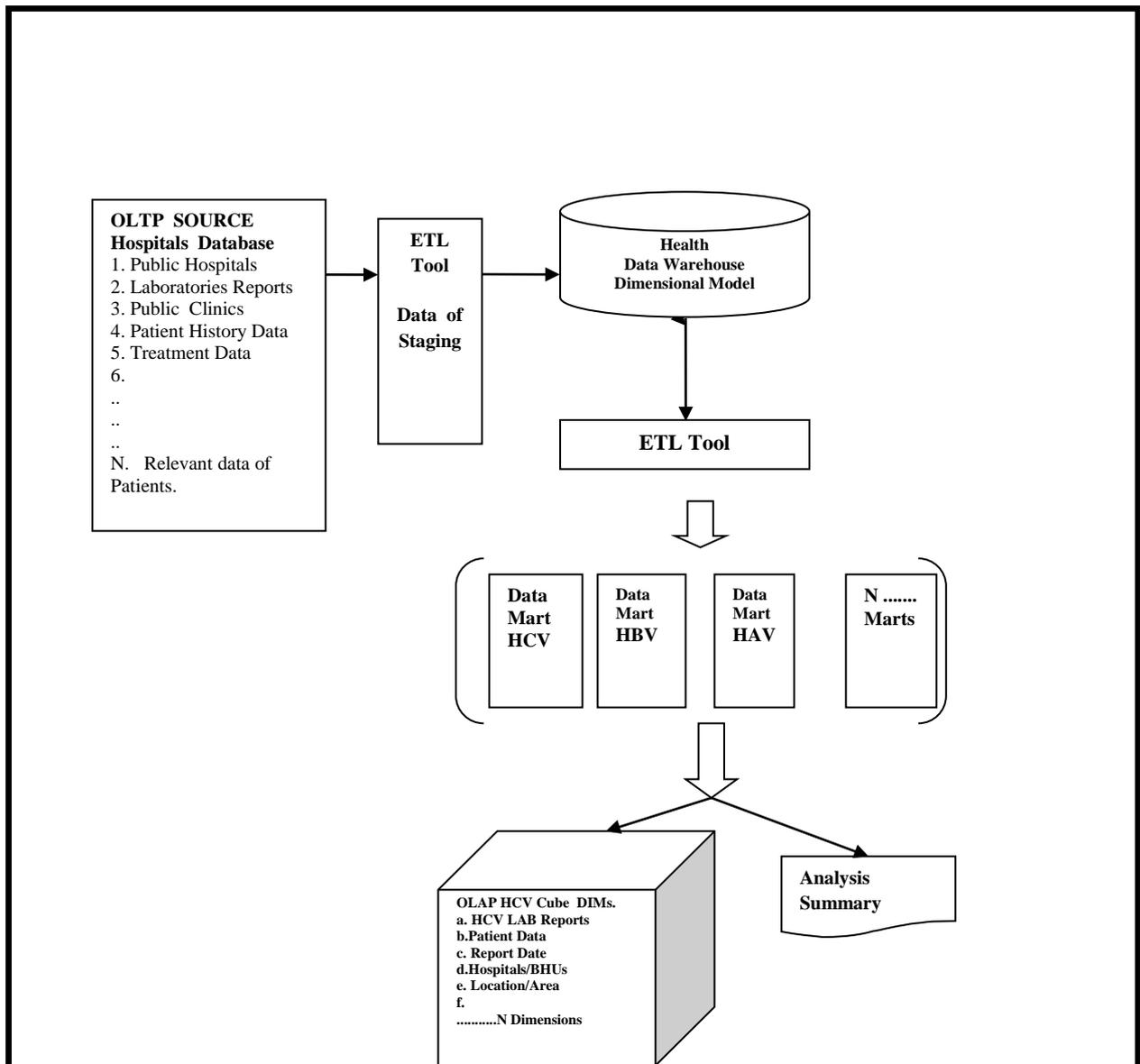


Figure 2: proposed Health Data Warehouse Model public Health sector.

5. Conclusions

The aim of this research was to design a conceptual frame work of Data ware housing to use in health sector, which provides data available in a distributed environment , data warehouses equipped with proper tools and mechanisms to integrate data with online approach. Initially designing the conceptual frame work of Health Data ware House model will be in government health sector only. As the amount of data and the number of systems involved increases rapidly, efficient and accurate data integration approach is required to create a single version of truth for all users of Health Dataware house in the country for analysis and development. The resolve the issue of health data integration it must be initiate the moderated DBMS in the Government Hospitals first to build a Health Database system then data ware houses can be implemented, possibly as there are three phases(conceptual, logical, physical) for implementation of Data ware house, in this study we discussed only the conceptual frame work and model through all corners and propose a useful and suitable model for Health Data ware house in government health sector.

6. Future work and directions for implementation of Health Dataware House in the country

The following listed are the future works that can be carried out based on this research to add more practical value.

- DBMS must be established in all hospitals to get the patient data for Data ware house.
- Implementing a Data Warehouse based by proposed conceptual model , a logical model would also be drawn using Data Warehousing and Integration software.
- Populate the Data Warehouse using Extract, Transform and Load tools.
- Run query and report generation operations by the Data warehouse for Analysis and Reports.
- Measure and document the performance of the data ware house.
- Develop a Data Integration framework which can be customized and adjusted according to the changing needs of Government Health sector.

References

- [1] Lacey Benson, et al. "When is family history obtained? – Lack of timely documentation of family history among overweight and hypertensive paediatric patients" .Jul. 2010 (Full publication history). DOI: 10.1111/j.1440- 1754.2010.01798.x .
- [2] Goldstein et al. "Neural substrates of the interaction of emotional stimulus processing and motor inhibitory control: An emotional linguistic" , 2007.
Available:http://levylab.psych.psu.edu/publications/2007/neural-substrates-of-the-interaction-of-emotional-stimulus-processing-and-motor-inhibitory-control-an-emotional-linguistic-go-no-go-fmri-study/at_download/file.
- [3] *Knaup P.* ,et al. "Electronic patient records: moving from islands and bridges towards electronic

health records for continuity of care.” 2007. Available:

<https://www.ncbi.nlm.nih.gov/pubmed/17700902>.

- [4] Hersh W. “Health Care Information Technology: Progress and Barriers”. JAMA. 2004, 292(18):2273-2274. Available: <https://doi.org/10.1001/jama.292.18.2273>.
- [5] Sahama & Croll (2007 “article” A Data Warehouse Architecture for clinical Data warehousing”,2010),
- [6] Thomas Handler et al. “HIMSS Electronic Health Record Definitional Model, Version 1.0”, by HIMSS Electronic Health Record Committee.2003.
Available:<http://www.hl7.org/Library/Committees/ehr/archives/030616-EHR%20Definitional%20Model%20Version%201.doc>. [Mar.7,18]
- [7] Pearce & Haikerwal (2010. Available:
<https://books.google.com.pk/books?id=nN7XDQAAQBAJ&printsec=frontcover>[March.4,2018]
- [8] Pedersen & Jensen (1998): “SSDBM '98 Proceedings of the 10th International Conference on Scientific and Statistical Database Management.”. July 01 - 03, 1998 IEEE Computer Society Washington, DC, USA ©1998. Page 43-52 .available: ISBN:0-8186-8575-1 doi>10.1109/SSDM.1998.688110.
- [9] Siliya Nugawela “Dataware Housing Model for integrating fragmented Electronic Health Records from Desperate and Heterogeneous Clinical Data Stores” Jun.2013.
- [10] Inmon W. H., Building the Data Warehouse. , New York: Wiley Computer Publishing, 2002. Available: <https://leseprobe.buch.de/images-adb/e2/e7/e2e7f926-314c-4e83-86a1-e481fafaf580.pdf>
- [11] Başaran, B. “A Comparison of Data Warehouse Design Models.” MS. Thesis. Computer Engineering, Atilim University, Turkey. 2005.
- [12] Martyn T. “Reconsidering Multi-Dimensional Schemas”, SIGMOD Record, Vol. 33, No. 1. Mar.2004.Available: https://sigmodrecord.org/publications/sigmodRecord/0403/B6.Martyn_6page.pdf.
- [13] Kerkri E. M., et al. “An approach for integrating heterogeneous information sources in a medical data warehouse”. Journal of Medical Systems, 2001, 25(3), pp.167- 176. Available: doi: 10.1023/A:1010728915998.
- [14] S. Aziz and M. H. Rao. “Existing Record Keeping System in Government Teaching Hospitals of Karachi”. DOW. Karachi, funded by Govt. of Pakistan. Apr.2002. Available: http://www.jpma.org.pk/full_article_text.php?article_id=2202

[15] US Department of Health; The World Health Organization. Available:

<https://books.google.com.pk/books?id=7W7QAQAAQBAJ&pg=PA65&dq=US+Department+of+Health;+The+World+Health+Organization&hl=en&sa=X&ved=0ahUKEwjOtYm3v9PZAhVNsKQKHWZqByoQ6AEIJTAA>.

[16] National Institute of Population Studies (NIPS) ,Pakistan, 2006 .\

Available:http://www.nips.org.pk/abstract_files/Priliminary%20Report%20Final.pdf. [Mar. 4, 2018]

[17] Altaf Bosan, et al. “ A review of hepatitis viral infections in Pakistan” 2010.

[18] Butt AS, et al. “Hepatocellular carcinoma in Native South Asian Pakistani Population; Trends, linicopathological characteristics & differences in Viral Marker Negative & Viral-HCC (A 10 year Experience)”,. 2011.

[19] Available: <http://mldata.org/repository/data/download/matlab/datasets-uci-hepatitis>.