Context-Aware Healthcare Dataset - A Case Study from Pakistan

Shahid Mahmud

Faculty of Engineering and
Computing
Coventry University,
United Kingdom
Email: mahmuds4@uni.coventry.ac.uk

Rahat Iqbal

Faculty of Engineering and
Computing
Coventry University,
United Kingdom
Email: r.iqbal@coventry.ac.uk

Faiyaz Doctor

Faculty of Engineering and
Computing
Coventry University,
United Kingdom
Email: faiyaz.doctor@coventry.ac.uk

Abstract—**This paper presents a context-aware healthcare dataset that has been designed to understand and monitor the health-shocks in Pakistan. Based on the socio-economic, cultural, and geographic norms, a user study based on questionnaire comprising of** 47 **features was carried out. In total,** $1,000$ **households belonging to** 29 **villages in rural areas participated in this user study. The purpose of this research is to monitor health-shocks in a community using data visualization and predictive modelling. We envisage that this study will provide insight into the relationships between socio-economic, demographic, and geographical conditions impacting health issues.**

Keywords–Context-aware; socio-economic; cultural; geographical; data visualization; and predictive modelling.

I. INTRODUCTION

Generally, healthcare systems are evaluated based on three main factors: quality, cost, and accessibility to health-care, known as "The Iron Triangle of Health-Care" as shown in Fig. 1. In case of an effective health-care system, there should

Efficiency/Cost Containment
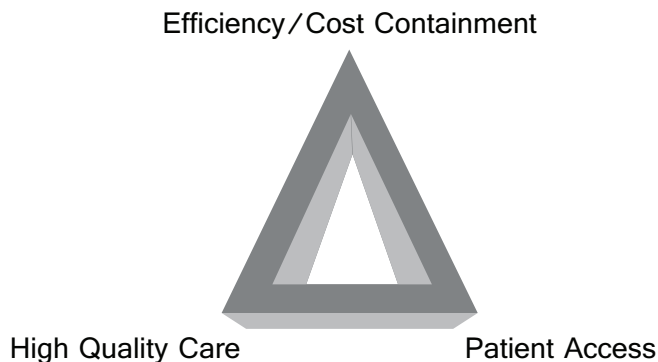
High Quality Care   Patient Access

Figure 1. The Iron Triangle of Health-Care [1].

be a balance between all the three components, i.e., the iron triangle should be an equilateral triangle, with each angle of $60°$ [1]. However, in practice, any effective health-care system can only optimize two of the three factors. For instance, to achieve higher access and quality, its associated cost will increase [2]. Furthermore, these factors highly depend on the socio-economic, geographic, and cultural norms. Especially, in order to understand the health-shocks situation of any third world country, there is a need to understand socio-economic, geographic, and cultural norms of that origin. By health-shocks, we mean critical illness of families, principle bread

winner and its socio-economic after-effects on individual, family, society and various governance levels [3].

In this regard, a lot of research work has been done to understand the reasons and effects of health shocks in the developed and developing countries [4]–[9]. In [7], different socio-economic factors and their impacts were studied. It was reported in [7] that 63.8% of health expenditure was out-of-pocket, i.e., from the pocket of patient, which resulted in financial losses. In [10], health related "hardship financing" for poor households in an Indian town Orissa was studied. The authors investigated factors influencing the risk of hardship financing with the use of a logistic regression. It was observed that in rural areas, most of the households were facing financial hardships due to indirect and/or long-term costs of health-care. In Orissa, 80% of spending on health-care was out-of-pocket of the households for which they either borrow money at higher interest or sell their assets.

From the various studies [10]–[12], it is quite evident that the unpredictable timing of health issues and immediate need for large funds for health-care in addition to the distance to health facilities could increase the risk of hardship financing.

In this paper, we have tried to understand the health-care system of Pakistan and how the socio-economic, geographical and cultural norms are affecting the health of almost 200 million Pakistanis, especially those who belong to rural and tribal areas. For instance, women in rural and tribal areas of Pakistan are not allowed by their men to consult a male doctor during pregnancy which results in higher infant mortality rates (IMR) and maternal mortality rates (MMR). In Pakistan, IMR which is a count of the number of infants that die before their first birthday in every thousand infants was 80 at the start of this century. Table I shows the IMR of Pakistan in comparison to other countries. Currently, there are only 25 countries that have a higher IMR than Pakistan.

Another useful measure for assessing children's health is their weight. Experts have figured out a scale that lists out the appropriate weight for healthy children at any given age. In 2001, percentage of underweight children who were less than five years old was 32% in Pakistan. During the millennium development goals, government of Pakistan has vowed to reduce it to 20%, by 2015. In developed countries, such as Japan, this percentage is less than one.

Similarly, MMR which is considered as a basic measure for assessing the health of the mothers in any given region was 490 in Pakistan during 1990. Another important factor that is directly related to IMR and MMR is "appropriate pregnancy

TABLE I. IMR of different countries [15]

Countries	1990	2010
Sweden	6	2
England	8	5
Malaysia	15	5
United States	9	7
Turkey	66	14
Sri Lanka	26	14
Saudi Arabia	36	15
China	38	16
Iran	50	22
India	81	48
Pakistan	96	69

spacing". Generally, it is recommended to maintain a 2.5 to 3 years gap between pregnancies which is vital for the health of both mother and child. However, such spacing is only possible with proper awareness, equipment and its availability, and birth control. According to [14], around 27% of the Pakistani couples who preferred to use some sort of birth control did not find it available in their local region.

II. Anatomy of Dataset

To understand the health shocks and its causes, we collected a dataset of $1,000$ households from the district Haripur with the help of Begum Mahmuda Welfare Trust hospital (BMWT). Haripur district is in the Hazara region of Khyber Pakhtunkhwa province of Pakistan. It is located in a hilly plain area at an altitude of around 610 meters above sea level with an estimated population of a million in 2009 [13]. In district of Haripur, there are about 39 hospitals and basic health centers, and 10 dispensaries. Furthermore, there is only one bed for every $1,516$ people [13]. In contrast to Haripur, there is one bed for every 100 people in the developed countries of the world. Moreover, in district of Haripur, IMR is 66 whereas overall MMR in the province of Khyber Pakhtunkhwa is 275.

Based on "patients to bed ratio", Pakistan is ranked at 178 out of 194, internationally. Furthermore, in comparison to developed countries where there is at least one doctor for every 712 people, Pakistan has only one certified doctor for every $1,230$ people.

A. Survey Features

In order to find the ground realities of health-care facilities in rural areas, we have collected the information about age, marital status, sex of the household head, their involvement in the labour force, education of children, financial and water resources, access to health facilities, schools and clean water, effects of climatic changes, effects of shortage of basic facilities like fuel, food, money for treatment of illness or fertilizer for crops, and waste disposal trends.

B. Ethical Considerations

Based on the cultural norms of this region, five ethical concerns were short-listed that we wanted to confirm while conducting the study. These six ethical concerns includes: voluntary participation, no harm to respondents, anonymity and confidentiality, identifying purpose, sponsor, analysis and reporting. Throughout the survey, these guidelines were followed strictly.

C. Survey Questionnaire

A printed questionnaire was used to obtain the survey information from the responders directly. Here, we preferred questionnaire over the interviews as questionnaires are more systematic, less prone to personal biases and transcription errors. Moreover, it offers more comfort to the responders as they can fill it in their own private settings with discretion. In our study, questionnaires were in national language of Pakistan, i.e., Urdu, in order to enable the villagers who does not understand English to be able to answer the questions. Furthermore, assistance was also provided to the participants who could not read or write. The questionnaire was divided into two sections, i.e., section A and section B.

Section A aimed at gaining demographic data such as age, level of education, income and gender whereas Section B is more concentrated on the living standard of the participants and effects of health shocks on their families. Furthermore, as the targeted population was too large to survey, so accidental sampling was used in identifying the participants. Here, "self-report" was used a responding mechanism where people voluntarily choose to respond to series of questions posed by investigators and are allowed to skip as many questions as they like.

D. Quality Assurance Measures

Different quality measures were also adopted from responder's perspective, i.e., understandability, comprehensiveness, and acceptability of the survey forms. We have used quite a few methods for quality assurance, including *sample testing* by asking few expert interviewers and responders to fill out the survey form and incorporating their suggestions, cognitive testing, by interviewing the responders to visualize the questions and reiterate them in their own words, behavioural testing by altering the questions and measuring the difference in responder's answers to find out how the wording in questionnaire will affect the overall answer of the responders, special probing, by explaining the intent of the questions to the responders in local language, so the responders would not take an infinite amount of time to answer the questions, experts opinion, by sharing the questionnaire with experts for their valuable feedback and suggestions, compare and contrast, by measuring the questionnaire against pre-existing surveys and their responses in order to see what value additions can be made, what best practices can be used, and what errors can be avoided.

E. Pilot Study

In the pilot study, a multidimensional survey questionnaire was distributed among 300 families living in the proximity of Haripur district. During the survey, geo-coordinates of 29 villages of district Haripur which participated in the survey were also noted as shown in Fig. 2. One of the primary objectives of the pilot study was to refine the survey questionnaire based on its feedback. During the pilot study, some families came to the hospital and/or local clinics to fill the questionnaire while others were contacted at their homes. The questionnaire was given to the household heads who were older than 21 years with one or more family members living with them, who were mentally stable, and were willing to participate in the survey. Furthermore, there was no race, religion, and gender discrimination during the survey activity.
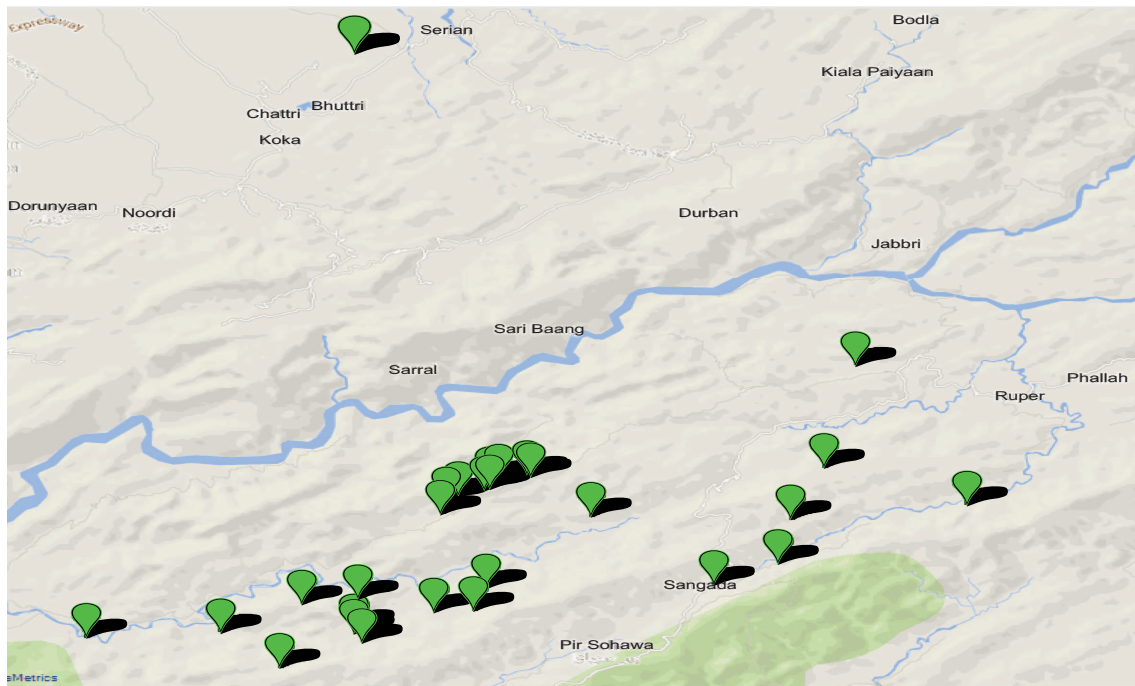
Figure 2. Geo-coordinates of the villages who took part in the survey.

During this pilot study, we also convinced the village elders to actively participate in the survey and to provide us a good insight about available health facilities in their villages. With the active participation of the village elders, we met number of families and explained them the purpose of this survey and the changes that it can bring in their daily lives. As a result, 300 households participated in the pilot study. We also made sure to maintain the confidentiality and integrity of the households by masking their names in the dataset.

As a result of the pilot study, we have found numerous information that have helped us to refine the survey, for example, most of the subjects have a hard time getting clean water or do not have a proper toilet facility. Almost all participants are single with at least two adults living with them. Students from participating families have to travel for at least 20 minutes to get to school. Minor illnesses occurred at least twice a year while major diseases occurred 3-5 times a year in most of the households. Most families fall short of money if anyone in the family falls ill or has an injury. Stone and mortar is mostly used as the basic construction material of external walls while ceilings are mostly made of thick wood. People mostly don't have any toilet facility at home. Waste food, water and garbage are mostly disposed off near homes. Almost every time during the year people have to rely on an irrigation canal for water source. Participants rely mostly on land and use it for agriculture or livestock. Most common hardships faced are loss of job or losing a house.

Furthermore, based on the feedback, we added some other information in the survey that affects the overall health-care system in Pakistan, namely: distance to the health units, number of basic health units, costs associated with travelling, accessible routes to the health facilities, vaccination, transportation, sewerage system, awareness, and water resources, just to name a few. Fig. 3 shows the modified iron triangle

that fits well to the socio-economic, geographic and cultural norms of our region.
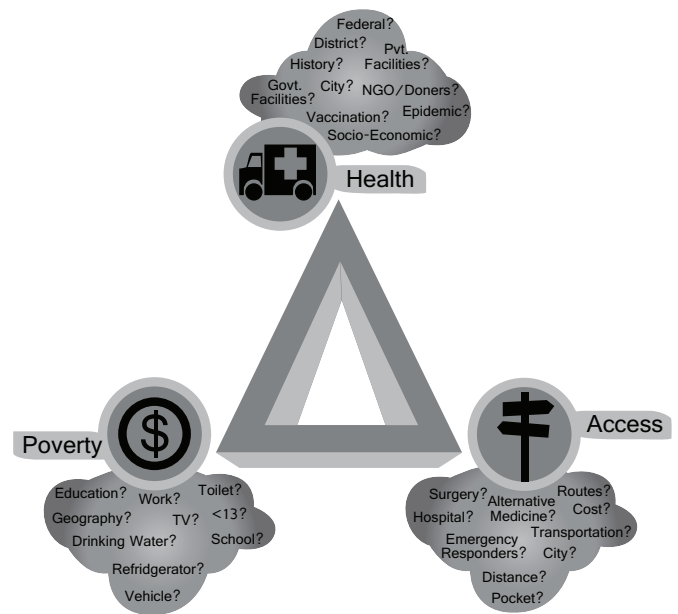

Figure 3. Factors affecting the Health-Care system of Pakistan.

F. Data Collection

At the end of the pilot study, all the changes were incorporated in the questionnaire. During the survey, we followed a structured approach in which all the responders were given the same possible choices and all questions were presented to the responders in the same order, i.e., instructions and

explanations were fixed. The questionnaire was filled by $1,000$ households. The head of the household was given a brief account of the research and its importance and the support of the administrator. Finally, we provided a telephone number to the village elder for anyone with questions or who may need assistance in completing the questionnaire at home. All participants were provided a comfortable environment and privacy. The questions were in easy to understand language. In case of unanswered questions, questionnaire was brought back to the participants to know the reason for not answering the questions. The reason was then noted on the questionnaire next to that question. In total, BMWT dataset comprises of $1,000$ households and for each household there are 47 features that plays a vital role in the health-care system of Pakistan. Table II shows the feature set of BMWT dataset.

TABLE II. Features of BMWT Dataset

Sr. No.	Features
1	Subject ID
2	Contact Number
3	Gender
4	ID Card No.
5	Marital Status
6	Age
7	Tehsil
8	Union Council
9	Village
10	GPS Coordinates
11	Adults living in house for more than 9 months in a year (Female)
12	Adults living in house for more than 9 months in a year (Male)
13	Adults earning
14	Distance of schools in Kilometers
15	Distance of schools in Minutes
16	Frequency of minor disease/year
17	Frequency of severe disease/year
18	Distance of basic health unit in Kilometers
19	Distance of basic health unit in Minutes
20	Distance of Hospital in Kilometers
21	Distance of Hospital in Minutes
22	Mid Wife During Birth
23	Distance of Vaccination Center
24	Polio Drops
25	Fatalities During Birth
26	Nature of Walls of House
27	Nature of Ceiling of House
28	Resistance of House against Severe Weather
29	Toilet Facility
30	Disposing off of food
31	Disposing of garbage
32	Disposing off of water
33	Dental Hygiene
34	General Hygiene
35	Water Source (most of the year)
36	Water Source (in dry weather)
37	Time Duration for collecting water for one day
38	Agricultural Land (in canals)
39	Expenses of Manure for Land
40	Domestic animals - Buffaloes/Cows
41	Domestic animals - Goats
42	Ownership of Land
43	Expected Problems (1st, 2nd and 3rd preference wise)
44	Solutions to expected problems
45	Duration for reconstruction of House in case of destruction (in months)
46	Shortage of food
47	Debts

III. DATA VISUALIZATION

Data visualization has been done using the purpose built "HexChange" tool that was developed in C# (We are working on its web-version which will be made publicly available). In BMWT dataset, 47 features of $1,000$ households belonging

to 29 villages of Haripur district were captured as shown in Fig. 4. Here, for the open-ended questions, we opted for the quantitative content analysis which is a formal, systematic, and objective process used in describing and testing the relationship and their causal interactive effects among variables.

Here, it is worth mentioning that all the features mentioned in Table II are highly dependent on each other. For instance, there is a positive relationship between distance to basic health units (BHUs), percentage of debts, toilet facilities, and frequency of major illnesses. Due to debts and distance to BHUs, minor illness turns into major.

Table III and Table IV shows the distance to BHUs and hospitals, respectively. It is clear from Table III that in Barkot, on average, each patient needs to cover a distance of 12.81 kilometres in order to reach a BHU with a standard deviation (std.) of 9.91 kilometers. Similar results can be seen in Ta-

TABLE III. Distance to BHUs in KMs

Union Councils	Min.	Max.	Mean	Std.
Barkot	0	34	12.81	9.91
Jabri	2	20	8.6	5.55
Musalimabad	9	18	15.27	2.63
Muslimabad	0	34	10.94	7.23
Najafpur	3	30	8.26	4.07

TABLE IV. Distance to Hospitals in KMs

Union Councils	Min.	Max.	Mean	Std.
Barkot	0	51	24.01	7.51
Jabri	10	34	27.21	5.56
Musalimabad	14	34	20.53	5.15
Muslimabad	0	36	25.62	7.92
Najafpur	10	120	18.23	11.18

ble IV, where each patient on average travels 24.01 and 25.62 kilometers from Barkot and Muslimabad to reach hospitals.

In district of Haripur, mean poverty score is 29.94 [16]. This poverty score is reflected in Table 5, where 49.55% of responders from Barkot are under the debt of more than 500,000 PKR.

TABLE V. Majority of people are in debt with moderate to large sums of money

Union Councils	Don't Know	No Debt	1K to 200K	200K to 500K	>500K
Barkot	5.04	4.75	15.13	25.52	49.55
Jabri	6.35	1.59	52.38	39.68	0.00
Musalimabad	40	6.67	46.67	6.67	0.00
Muslimabad	1.05	12.87	11.60	29.75	44.73
Najafpur	4.72	8.49	21.70	65.09	0.00

Due to higher distance of BHUs and hospitals in Barkot, people have to pay much more in order to reach the health facility. Here, cost of travelling is more or equal to the cost of treatment which results in higher debts and severe illness. Table 6 shows the frequency of major/severe diseases. Here, major/severe disease is defined as an injury which requires two or more days of bed rest or hospital admission. It also includes disability.

Table IV-VI highlight the relationship between distance to BHUs and hospitals, debts, and major illness which supports our observation that socio-economic, geographical and cultural
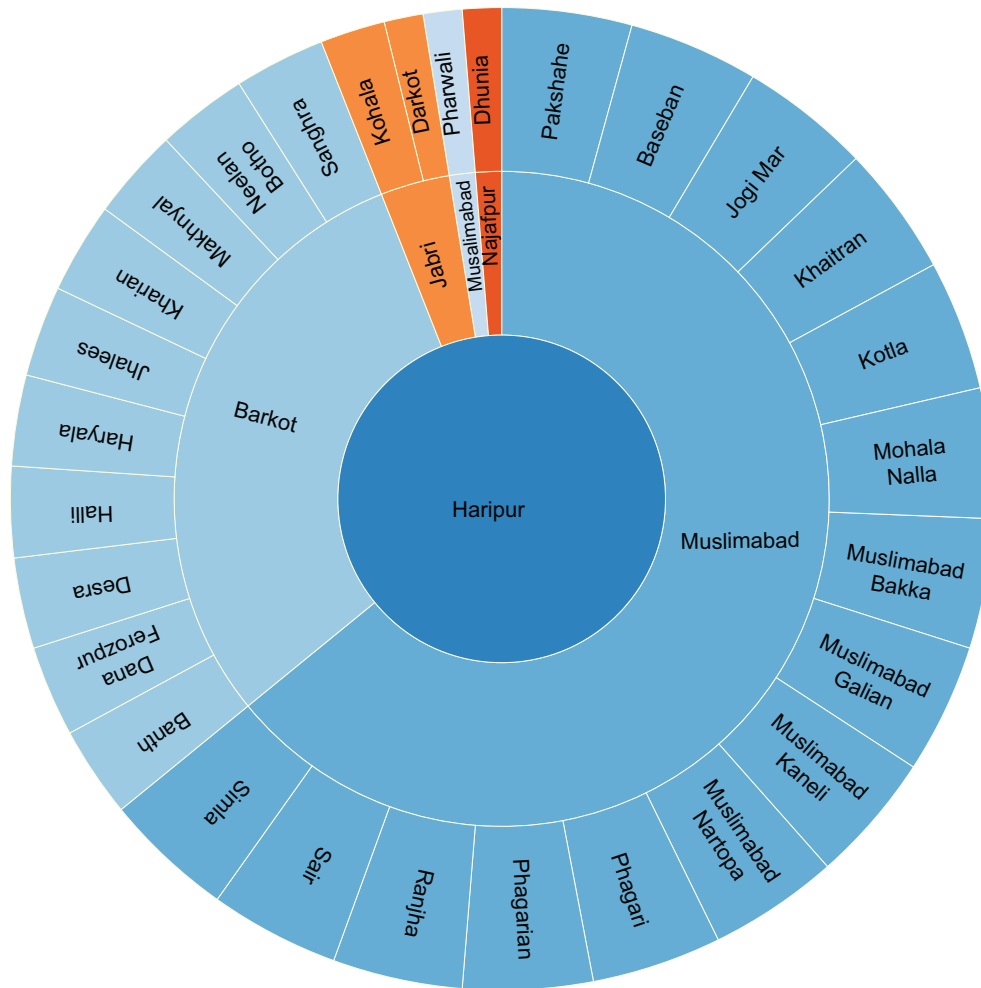
Figure 4. Villages that participated in the survey. Inner most circle represents the district Haripur whereas the outer circle represents the union councils and the outer most circle represents their corresponding villages.

TABLE VI. Frequency of Major Disease in Percentage

Union Councils	Never	Yearly	Bi-Monthly	Monthly	Bi-Weekly	Weekly	Daily
Barkot	5.06	13.99	7.44	20.83	14.58	24.7	13.39
Jabri	1.59	17.46	4.76	11.11	1.59	50.79	12.7
Musalimabad	13.33	20	6.67	13.33	6.67	20	20
Muslimabad	9.92	20.46	11.39	20.89	10.13	12.66	14.56
Najafpur	12.26	10.38	6.6	19.81	14.15	25.47	11.32

norms highly affects the health-shocks, especially in the rural and tribal areas.

Furthermore, it is interesting to see the houses with toilet facilities have higher rate of minor and major diseases in comparison to houses with no toilet facility as shown in Fig. 5. One of the main reasons was access to water resources and poor sewerage system. Here, it is worth mentioning that the time required for household to collect water for one day usage is almost 4 hours as in some cases, it takes a women 1.5 to 2 hours to reach the source. Same amount of time is required to carry that water back home. Especially, in case of a family with two to three children, it requires more than 5 or 6 buckets of water at least for a day.

IV. CONCLUSIONS

Currently, there is no publicly available dataset that can help to understand and monitor the health-shocks in Pakistan. Such kind of surveys and datasets can be helpful to government - who would use this dataset and resulting analysis to form policies, to general practitioners and NGOs, in order to start community based health programs. This dataset is our first initiative to analyse and understand the health-care system and health-shocks, especially in rural and tribal areas of Pakistan (For those who are interested in BMWT dataset, please contact: mahmuds4@uni.coventry.ac.uk). Our proposed future work is to apply machine learning techniques to an extended dataset sampled from a larger population to develop a predictive model of health-shocks that forms part of a framework which
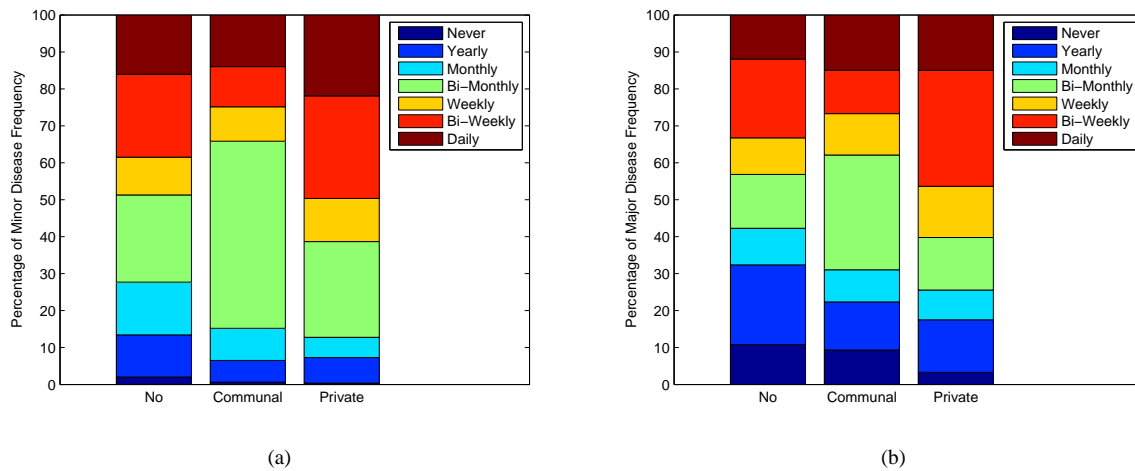
Figure 5. (a) Frequency of minor disease versus toilet facility. Here, communal represents a toilet shared by more than 3 people. b) Frequency of major disease versus toilet facility.

uniquely accounts for the cultural and traditional norms of this part of the world. The aim of this data intelligence driven framework will be to provide tailored and informed health-care analysis to stakeholders towards facilitating a national agenda of health-care reform.

REFERENCES

[1] K. William, "Medicine's Dilemmas," in *New Haven, CT: Yale University Press, 1994.*

[2] L. R. Burns, "A System Perspective on India's Health-Care Industry," in *global economy, health-care reform, India, World Development, July, 2014.*

[3] S. Mahmud, R. Iqbal, and F. Doctor, "An Integrated Framework For The Prediction Of Health Shocks," in *Proceedings of The 2nd International Conference on Applied Information and Communications Technology (ICAICT), April, 2014.*

[4] R. M. Townsend, "Risk and insurance in village India," in *Econometrica, vol. 62 (3), pp. 539591, 1994.*

[5] A. Kochar, "Explaining household vulnerability to idiosyncratic income shocks," in *American Economic Review, vol. 85 (2), pp. 159-164, 1995.*

[6] E. Skoufias and A. Quisumbing, "Consumption Insurance and Vulnerability to Poverty: A Synthesis of the Evidence from Bangladesh, Ethiopia, Mali, Mexico and Russia," in *The European Journal of Development Research, vol. 17 (2), pp. 24-58, 2005.*

[7] M. R. Howlader, "Analysing the socio-demographic variables impact on health status of Bangladesh," in *social science research network, 2013.*

[8] R. Iqbal, N. Shah, A. James, and J. Duursma, "ARREST: From work practices to redesign for usability," in *Journal of Expert Systems with Applications, vol. 38 (2), pp. 1182-1192, 2011.*

[9] F. Doctor, R. Iqbal, and R. Naguib, "Fuzzy Ambient Intelligent Agents Approach for Monitoring Disease Progression of Dementia Patients," in *Journal of Ambient Intelligence and Humanized Computing (AIHC), vol. 5 (1), pp. 147-158, 2014.*

[10] E. Binnendijk, R. Koren, and D. M. Dror, "Hardship financing of health-care among rural poor in Orissa, India," in *BMC Health Services Research, 2012.*

[11] Robert Wood Johnson foundation, "Changes in health-care financing and organization," in *2009.*

[12] R. Victor, F. Jeffrey, and E. Harris, "Review: Who Shall Live? Health, Economics, and Social choice," in *The Bell Journal of Economics, vol. 7 (1), pp. 340-343, 1976.*

[13] Government of Pakistan, "Khyber Pakhtunkhwa Health Sector Situation Analysis," in *December, 2010.*

[14] W. Hameed, S. K. Azmat, M. Bilgrami, and M. Ishaqe, "Determining the factors associated with Unmet need for family planning: a cross-sectional survey in 49 districts of Pakistan," in *PJPH, vol. 1 (1), 2011.*

[15] –, "UN Inter-agency Group for Child Mortality Estimation (UNICEF, WHO, World Bank, UN DESA Population Division)," *2014.*

[16] Durr-e-Nayyab, "Population Dynamics in Pakistan: An Analysis of the BISP Poverty Score Survey," *2010.*