An Efficient REDCap Based Data Collection Platform for the Primary Immune Thrombocytopenia and Its Analysis Over the Conventional Approaches

Nadeem Ahmed[†], Md. Shamsujjoha*, Md. Nawab Yousuf Ali[‡], William Marsh[┼]

[†]Department of Computer Science and Engineering, University of Asia Pacific, Dhaka-1215, Bangladesh.
[*‡]Department of Computer Science and Engineering, East West University, Dhaka-1212, Bangladesh.
[┼]School of Electrical Engineering and Computer Science, Queen Mary, University of London, London, England.
Email: [†]nadeem11@gmail.com, *dishacse@yahoo.com, [‡]nawab@ewubd.edu, [┼]d.w.r.marsh@qmul.ac.uk

***Abstract*— Primary Immune Thrombocytopenia (ITP) is a rare chronic disease. The registry of ITP contains all the necessary information for its research and analysis. The registry is thoroughly dependent on its underneath data collection platform. Generally, data collection for the ITP conducted on a routine basis and over a long period of time. However, most of the existing web based data collection approaches compromise the data security, data backup, flexibility and scalability. The REDCap based system can overcome these weaknesses as it has easier migration, validation, extraction and evolution processes. Thus, this paper develops a REDCap based data collection platform for the ITP registry. Then a detail comparative analysis between the two systems has been presented.**

Keywords—Chronic Disease, Data Collection Platform, Primary Immune Thrombocytopenia, REDCap, Registry, Research.

I. INTRODUCTION

In health studies, smooth and timely data collection is very important. It ensures the development of novel medical treatment and provides necessary background for the research studies specifically in case of chronic disease such as ITP [1]. ITP is a rare autoimmune disorder that can lead to easy or excessive bruising and bleeding [2]. It can be a long-term concern for the patients and may affect even after a symptom-free period as discussed in [3, 4]. The data collection of ITP occurs yearly or bi-yearly on a routine basis and over a period. Hence, the research and analysis of ITP fits under longitudinal study. Data collection for the ITP researches is done through the online surveys and web based data entries [5]. Generally, research team employs a third party solution provider to create a suitable web system for this purpose. The web based data collection scheme inherently compromises security, flexibility, data backup, availability, and scalability. In addition, these systems provide an inconvenient platform for end user and leave few laborious tasks for the administrator. In this paper, we show that the **R**esearch **E**lectronic **D**ata **Cap**ture (REDCap) tool can be a potential alternative to resolve the mentioned problems especially for the ITP registry. The ITP registry is developed and maintain by the Barts and London School of Medicine and Dentistry and The Royal London Hospital. It collects data from the participating sites (using web based live stream). These data assist understanding of disease and its effective treatment strategies [6].

The existing registry system has both technical and organizational weakness. Organizational issues are beyond scope of this paper. In technical aspect, the registry faces data validation problem, for example drug overdose should generate a warning message. The format of output data is presented in MS Access database and cannot be fed directly to existing tool such as Stata. The research team uses this tool for data analysis. These data initially goes through a manual cleaning process, which is time consuming and also laborious. On other hand, output in MS Access format is not suitable for clinical research and analysis. In this context, REDCap is considered as a potential alternative tool to resolve these issues. REDCap is a secure web application for building and managing the online surveys and databases [7]. It is easily manageable and configurable. Moreover, the software and support are available from institutional partners without any additional charges. In these consequences, this paper developed an efficient REDCap based data collection platform for the ITP registry. Then detailed comparative analysis between the developed system and existing web based approaches has been presented. The rest of the paper is organized as follows: Sec. II discusses necessary background which is required to understand the proposed work. Then, a brief discussion on REDCap and data migration is shown in Sec. III. Finally, based on all these discussions a conclusion is drawn in Sec. IV.

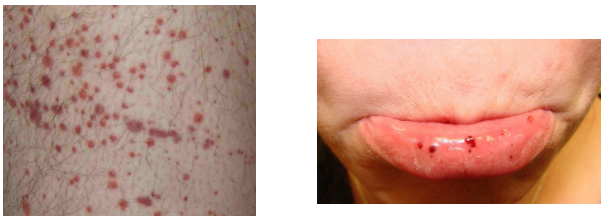
II. BACKGROUND STUDY

This section starts with the discussion about ITP disease. Then a brief introduction about the ITP registry is presented. The section ends with the discussion on REDCap.

A. ITP Disease: Diagnosis and Treatments

ITP is a bleeding disorder characterized by the low platelet count. There is no specific underlying cause for this disease although an auto immune mechanism is suspected. Regular and periodical check-up are needed when ITP turns to a chronic disease. Analysis of ITP is deeply depended on the data, which is collected in a long term. The diagnosis of ITP is done by a blood test known as full blood count. Usually a lower number of platelets are found in this case. Sometimes, physicians also suggest microscope testing because lower platelets might occur due to medication, lupus or viral infections [6, 8]. Sometimes sample from bone marrow is also taken for adults.

978-1-4673-9930-2/15/$31.00 ©2015 IEEE



(a) (b)

Figure 1. Brusing in the human body due to ITP
(a) Small bruise-like markings (b) Oral petechiae lower lip [9]

Treatments for ITP generally avoided for the non-adults. Even for low platelet counts treatment is usually avoided for children. If there is severe bleeding or bruising (shown in Fig. 1) treatments are provided. However, for adults, treatment can be either first-line treatment or second-line treatment, whose success rate depends on bleeding history, comorbidities, family history, consent etc. [9].

B. ITP Registry

ITP Registry (ITPR) is multi-centre rare disease registry maintained that collects basic demographics, clinical data and tissues. But from 2007, it also collects epidemiological data. A third party solution provider was appointed in 2010 to build a web-based and online database system for the registry. The registry also engages data linkage services provided by the Health and Social Care Information Centre (HSCIC) which complements the data collection from the medical records, soon after it realizes the importance of practitioners and request data from them. Currently there are fifty two centers in UK that upload patients data on the database [6, 10].

C. Brief Survey on REDCap

REDCap was first introduced at Vanderbilt university with the support from National Centre for Research Resources (NCRR) and National Institutes of Health (NIH) in 2004 [11]. Main goal was to develop a cost effective tool to facilitate the clinical research and support the institutional researchers [11]. REDCap can be considered as a web application that is fully developed and reliable for creating and mastering online surveys and databases. The REDCap has around two hundred thousand projects with over three hundred thousand users. REDCap projects are created one of the two ways:

i. Online method (online designer): using web browser

ii. Offline method (data dictionary): filling up MS Excel template, which can later be uploaded into web.

The REDCap software allows audit trails for tracking user activity and data manipulation. It also supports varieties of data export formats *e.g.*, MS Excel, PDF, SAS, Stata, etc. The software has scheduling modules, ad-hoc reporting tools, different language supports and easy installation process. REDCap has several releases. To the best of our knowledge, most recent version is 5.12.0. The global REDCap team is persistently working to develop more options and fulfill the user requirements [11].

REDCap is designed for the non-IT background people to create and manage their own projects. It is obligatory for a research team to have knowledge of designing, managing and maintaining the REDCap project as most of the cases REDCap project is created and managed by the research team. Although local consortium can support the research team but no one except the research team knows better about the research and its data. Thus, learning directly provides flexibility for future research and analysis. Learning resources are free of charge and available in REDCap website [11]. Local partners also publish training materials in their corresponding websites and arrange training for their users. The proposed project uses the materials of Institute for Clinical and Translation Science, University of IOWA [12]. These resources are good enough to create forms smartly and efficiently.

An institutional partner can use the REDCap software with a valid end-user license agreement with Vanderbilt University. According to the policy of participation rules, a dedicated IT team (preferably with a system administrator) with network infrastructure is required. Though, it is not difficult to support REDCap infrastructures and each consortium site is individually accountable for their system installation, maintenance and launching. REDCap can operate on different operating systems. A list of required hardware and additional software is given below:

i. Web server with PHP: PHP 5.1.2 and up (Apache, Microsoft IIS), cURL, MySQLi extension in PHP.

ii. MySQL database server (MySQL 5.0.0 and up) and MySQL client

iii. SMTP mail server and File server (optional)

All these components can share single machine but it is recommended to have the separate web, database and file server for performance as well as safety. Generally, it is suggested that both web server and MySQL should separately have 10GB space for smooth operation [6].

III. PROPOSED DATA COLLECTION PLATFORM

In this section, we describe the design and working procedure of the proposed data collection platform. Initially, the section describes the data migration part. Then the working procedure of the proposed system has been described.

A. Design Procedure: Data Migration in the Proposed REDCap Based System from the Existing System

Let assume, the patients data already exist. Then it needs to be transferred to the REDCap based system. An automated procedure is needed as it is infeasible to enter all these data manually. Thus, data migration is the most crucial design decision. It consists of several steps which are as follows: The initial step is to transfer the entire database. As mentioned earlier, ITP registry gets an MS-Access database as an output in the existing system. Therefore, we have decided to move the MS Access database to SQL server as the SQL server has several advantages such as faster data access and flexible environment for analyzing the data.

	A	B	C	D	E	F	G
1	patient_id	surname	forename	dob	reg_date	gender	ethnic
2							
3							
4							

Figure 2. Data structure for patient information

In the second step, the relationships between the tables and data flows are interpreted. We found that the MS Access database is an underlying version of the original database where patient identification number (*patient_id*) is considered as the primary key. In the existing system, all the tables in the database are directly or indirectly related with a patient identifier (*patient_id*). Thus, the data migration becomes much easier from the existing web based system to the proposed REDCap based system. An approximate structure of the patient database is shown in Fig. 2. Then the import tools such as Microsoft SQL Server and Business Intelligence Studio have been used for data migration. According to our design procedure, all the fields appear horizontally for more than one instrument. The proposed system has fifteen instruments and two thousand fields.

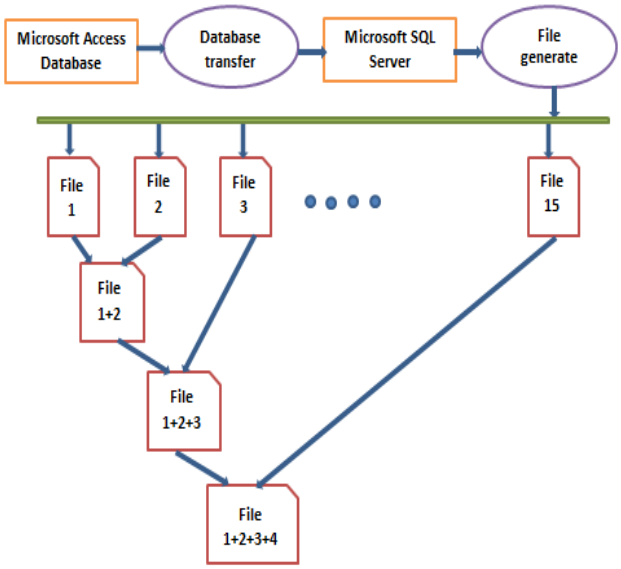


Figure 3. Process of data migration in the proposed system

In the final step, we write the data for each field (in a row for a single patient) in the database. The *patient_id* is kept as primary key to maintain the similarity with the existing approaches. It also allows performing *join* operation. Each instrument has one CSV and file hence we have fifteen CSV files. This design procedure is shown in Fig. 3. Initially *join* operation is performed between the files 1 and 2, then between the files (1, 2) with 3, then files (1, 2, 3) and 4, and so on. The final file contains all data that can automatically be imported to the proposed REDCap based system.

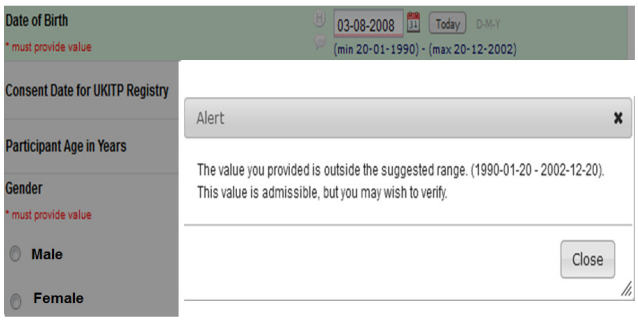
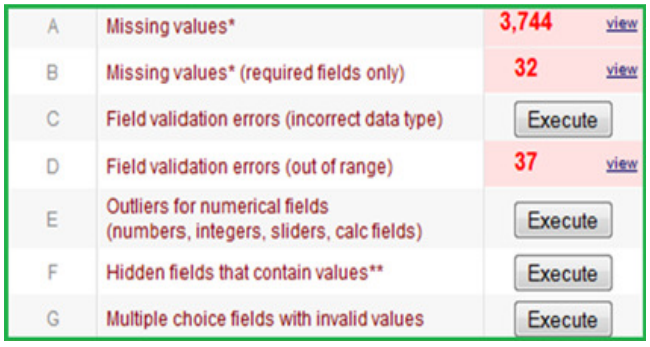


Figure 4. Process of data validation

B. *Woring Procedure of the the Proposed REDCap Based Sytem with Respect to Existing Web Based System*

1) Data Validation: The proposed REDCap based system has several built-in data validations options. These validation options are dynamic, for example the user can't enter future date as birth date and patient age must be over eighteen. In addition, statically minimum and maximum date in the validation option can also be entered for example, a warning window automatically poped up when the user enters beyond the statically set ranges. This is shown in Fig. 4. There is another built-in option named *data quality*. It provides seven built-in data checking (A to G) features. Fig. 5 shows the working procedure of the data checking in the proposed REDCap based data collection platform. From Fig. 5(a) we find that the extension can easily verify missing fields, meaningless data and violated data. In addition, the center/user responsible for entering wrong data can also be identified using this quality option. On the other hand Fig. 5(b) shows that the checking rule that can be adopted by the user such as age over sixty.



(a)

(b)

Figure 5. Data quality checking in the proposed platform (a) Built-in rules (b) User defined rule

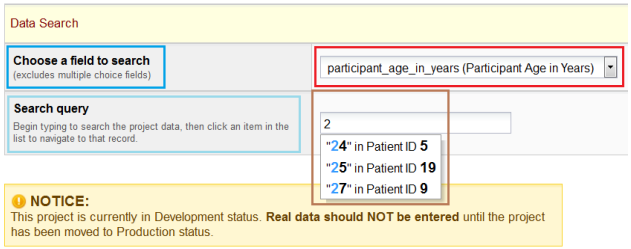


Figure 6. Working procedure of the search option

2) Search and Query: Fig. 6 shows the working procedure of the proposed REDCap based systems search option. It allows searching with only one field at a time. However, few cases demand searching with the combination of several fields which is unavailable in the REDCap based system.

3) Data Extraction, Formatting and Reporting: Proposed REDCap based system has built in option for data download. The option is named as Data Export Tool and available under Applications. In addition, downloaded data can be extracted in various formats such as Excel, PDF, SPSS, SAS, Stata etc. The extracted data are quite different in format with respect to access database of the existing system. In REDCap based system, all the fields in the forms have appeared as a long row. Some of the forms required conversions to make the forms as panel data, but the process is fine for the research team. Another advantage is date format in the outputs. Thus, data cleaning is not required with STATA or similar statistical tools. REDCap based system also provides two built-in applications that are Graphical Data View and Stats and Report Builder. These applications provide basic reporting options.

4) User Knowledge and Security: Any REDCap based project is solely designed, created and maintained by the research team who are responsible for sorting out the problems. The existing system for the ITP registry requests MDSAS (solution provider) for the change request, modification or update. However, the REDCap based system can change, modify or update by the user itself. Security gets the highest priority for any REDCap based system. Since, a single REDCap installation is designed to support many projects and hence, it is possible to provide VM server and actual server. Here, only nominated admin staff can access the server console for administrative purposes. IP blocking is also possible. The *ssh* access is blocked for all IP and is only possible via certain secure routes. The *https access* is allowed from all addresses, which allows global participation. However, it locks the account if there are five incorrect attempts in fifteen minutes. The CentOS is considered as the operating system for the proposed project. The project is pro-actively maintained with regular releases. Any potential security issues are patched quickly by the vendor.

5) Flexibility, Availability and Backup: The proposed REDCap based system is much more flexible than the existing system. It allows changes after production. There are varieties of options that can be added, implemented or even changed in the REDCap system. The system also provides flexibility through its backup procedures. It is maintained by Tivoli storage manager. Tape are back-up onsite and replicated to the secure remote site. The server is a VM and the underlying storage is RAID. The system is designed for multiple concurrent users and plenty of projects. The system is scalable and more web servers could be introduced behind the load balancer in theory. It also provides activity graphs to monitor load. The availability depends on the server, where the system was set-up. REDCap provides necessary supports in case of problems, but it depends on the agreement between IT department and research teams. Generally, research team is responsible for design, implementation and maintenance of any REDCap project. The supports are provided locally. In REDCap based project implementation, each centre can have different user roles with different user permissions. Creation of user roles and permissions is maintained and provided by local authority. In the proposed system, different user accessibility options are provided. The physical block diagram is shown in Fig. 7 [6].

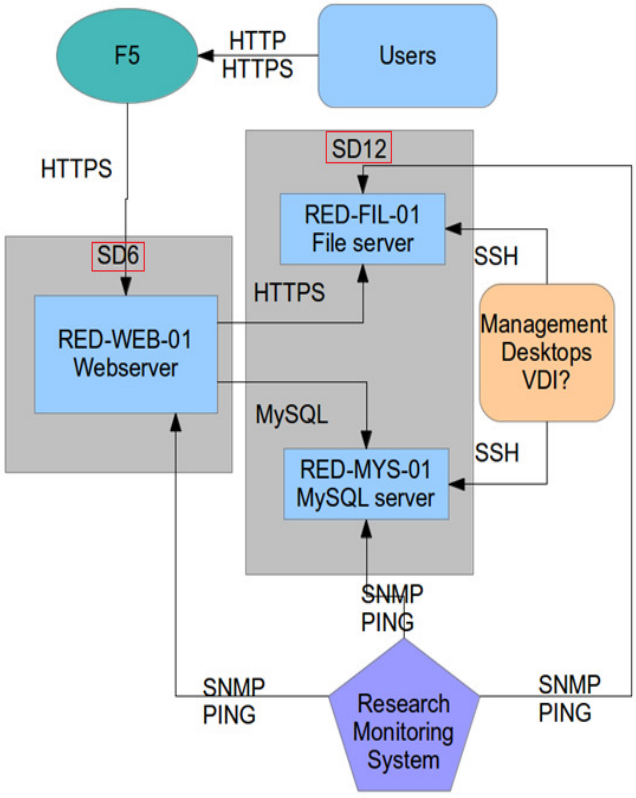


Figure 7. Architectural block diagram of the application [13]

C. Recommendation and Other Issues

Primary advantage of the REDCap based project development is its easy design process. Moreover, REDCap is customized software developed for non-IT people. On the other hand, the existing web based system is developed using ASP.NET with SQL Server as backend database, which definitely requires specific IT skills for its user to create forms and manage database. Moreover, the exisiting system requires third party solution provider for the system setup and support, whereas REDCap can be maintained by the research team themselves. This is one of the prime goals that has been achieved in this project. In REDCap system, research team has the full familiarity, awareness and knowledge of the project as well. Another prime advantage REDCap is its cost free access. An institutional partner can use the REDCap software with a valid end-user agreement with Vanderbilt University. On the other hand, the registry has to pay handsome amount of money for the annual fee of database maintenance and support to MDSAS. However, outlook is one of the minor limitations in the REDCap based system. It provides limited formatting options. All the fields in any form generally appeared vertically one after another. Here font, color, size, style and position can be changed. Complex outlook may result inconvenient platform for the user *e.g.*, users may find it difficult, while navigate through the system. On the other hand, every possible type of complex design and outlook is possible with conventional technologies. Another limitation of the REDCap based system is, every field needs to be created manually. This may take a lot of time and effort specially for the first time. As an example let assume that we need to enter data for two hundred bleeding events for a patient in the developed system. Then, we need to create individual fields for all two hundred bleeding events. This is laborious for the user, but it occurs as we want to minimize the cost. Thus, it perfectly matches the requirements of relatively small or medium research. As a tool, REDCap has several solid features. The format of data is absolutely suitable for analysis in Stata software. The API can play significant role in data validation and automated tasks as well.

IV. Conclusion

This paper presented an efficient REDCap based data collection platform for the ITP registry. Then an elaborate analysis between the REDCap based system and existing web based approaches has been shown. It has also been discussed that each system has its own advantages and neither of them is superior over another. However, the technical choice is dependent on the relative facts.

The paper also evidenced that the REDCap API can be a potential tool for accurate data validation and automated task for the future research works. It is suggested that the research team should continue with the existing system in parallel with the REDCap based system for a short time period (four to six months). This will make the user flexible in the new system and further performance analysis can also be accomplished.

V. References

[1] J. N. George., "Definition, diagnosis and treatment of immune thrombocytopenic purpura", Haematologica, vol.94, no.6, pp.759-762, 2009.

[2] A. K. Parekh, R. A. Goodman, C. Gordon, H. K. Koh, and The HHS Interagency Workgroup on Multiple Chronic Conditions, "Managing Multiple Chronic Conditions: A Strategic Framework for Improving Health Outcomes and Quality of Life", Public Health Reporters, vol.126, no.4, pp.460–471, 2011.

[3] C. J. Mann, "Observational research methods—Cohort studies, cross sectional studies, and case–control studies", African Journal of Emergency Medicine, vol.2, no.1, pp.38–46, 2012.

[4] Anand K. Parekh, MPH, Richard A. Goodman, Catherine Gordon, Howard K. Koh, and The HHS Interagency Workgroup on Multiple Chronic Conditions, "Managing Multiple Chronic Conditions: A Strategic Framework for Improving Health Outcomes and Quality of Life," Public Health Reporters, vol. 126, no. 4. pp. 460–471, 2011.

[5] C. W. Wesley, C. F. Alfonso, and T. K. Ricky, "A knowledge-based multimedia medical distributed database system", Information Systems, Scientific databases, vol.20, no.2, pp.75-96,1995.

[6] The ITP Support Association, [online] http://www.itpsupport.org.uk, (Last accessed Aug. 10, 2015).

[7] Paul A. Harris, Robert Taylor, Robert Thielke, Jonathon Payne, Nathaniel Gonzalez, and Jose G. Conde, "Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support", J. of Biomedical Informatics, vol. 42, no. 2, pp. 377-381, 2009.

[8] D. Provan, R. Stasi, A. C. Newland, V. S. Blanchette, P. Bolton-Maggs, J. B. Bussel, B. H. Chong, D. B. Cines, T. B. Gernsheimer, B. Godeau, J. Grainger, I. Greer, B. J. Hunt, P. A. Imbach, G. Lyons, R. McMillan, F. Rodeghiero, M. A. Sanz, M. Tarantino, S. Watson, J. Young, and D. J. Kuter, "International consensus report on the investigation and management of primary immune thrombocytopenia", Blood, vol. 115, no. 2, pp. 168-186, Jan 2010.

[9] Kenny T, and Willacy, "Understanding of Idiopathic Thrombocytopenic Purpura", [Online] http://patient.info/health/immune-thrombocytopenia-leaflet, (Last accessed Aug. 10, 2015).

[10] S. Akbayrama, F. Aktarb, C Akgünb, M. S. Bektaşb, H. Çaksenb, and A. F. Onera, "A case of immune thrombocytopenic purpura presenting with intracranial hemorrhage", Journal of Acute Disease, vol. 2, no. 3, pp. 250–251, 2013.

[11] REDCap, [Online]. http://www.project-redcap.Org, (Last accessed Aug. 10, 2015).

[12] REDCap documents, "Institute for Clinical and Translational Science, University of IOWA", [Online] https://www.icts.uiowa.edu/confluence/display/REDCapDocs/REDCap, (Last accessed Aug. 10, 2015).

[13] P. Butcher, "Platform for REDCap Request Document", REDCap Request, London: QueenMary, University of London, ITS Research Group. Report No.: Version: 1.4.1 Final Release, 2013.