Using R and Latex to Develop Automated Wait Time Report Generator for Nova Scotia Screening Program

By

Abdulhady Habash
B00530896
ab490448@dal.ca

Performed at
Nova Scotia Breast Screening Program
Halifax Shopping Centre
7001 Mumford Rd.
Unit 603L
Halifax, NS
B3L 2H8

In partial fulfillment of the requirements of the Master of Health Informatics Program, Dalhousie University

Report of Internship for the period May 3rd – August 15, 2011

Date Submitted: August 15, 2011

\

Table of Contents

List of Figures

Acknowledgements

Developing the AutoReport Generator software would not have been possible without the support of many people. I am grateful to Mr. Mohamed Abdolell, Associate Professor of Diagnostic Radiology & Division of Medical Education/Informatics, and to Mrs. Theresa Foley, Nova Scotia Breast Screening Program manager, both of whom were abundantly helpful and offered invaluable assistance, support and guidance. Special thanks also go to the workers at the NSBSP, especially to Mr. Ryan Duggan, Data Analyst/Project Coordinator and to Mr. Stephen Hankinson, Database Administrator. Finally, I wish to acknowledge the contributions of my colleague, Mr. Husam Al Qatami, whose hard work and valuable assistance helped bring this project to fruition.

Executive Summary

For women in Canada, breast cancer has gradually grown to become the most commonly occurring cancer type. Even though cases in the latter stages of the disease are hard to treat, the outcome for cases caught in the early stages is positive. Therefore, early detection is essential if the disease is to be successfully treated, and the most common and useful way to detect breast cancer at the early stages is screening tests. However, due in part to the increase in the occurrence of the disease as well as the growing popularity of screening programs, wait times for screening and diagnostic procedures have become almost unmanageable. This unfortunate turn of events has prompted researchers and statisticians to explore and develop options to improve wait times. The current most popular method of calculating wait times is through manual methods, but these are both substantially time- and cost-consuming. Other current methods utilize older-style databases that result in erroneous calculations. In addressing this urgent issue of improving the method of calculating wait times, this paper proposes the following: using a database structure that has a more appropriate design, and incorporating standardized methodologies in statistics and graphics. In enabling this solution, three open-source softwares were chosen for their easy accessibility, effectiveness and migration between platforms. The three applications used in this project were as follows: Caisis (a database management system); R (a statistical analysis program); and Latex (a document preparation system). Through the use of Caisis, R and Latex, an AutoReport Generator was developed that was able to create wait time reports quickly and cost-effectively. It is hoped that using the AutoReport Generator will help in the fight against breast cancer by creating more effective wait time management.

Organizational Profile

The Canadian breast cancer program has several attributes, one of which is the Nova Scotia Breast Screening Program (NSBSP) [1]. NSBSP is a provincial screening program founded in 1991 [2]. At that time, it was the fifth program in Canada to offer comprehensive breast screening services. The main office of this program is located in Halifax, the capital city of Nova Scotia, at the Halifax Shopping Centre, 7001 Mumford Road. The main goals of this program are:

- Reducing mortality related to breast cancer among women between 50 and 69 years of age by at least 30 percent in the province of Nova Scotia through mammography [1].
- “Developing standards, guidelines and policies to support a decentralized model of breast screening at multiple sites throughout Nova Scotia” [2].
- “Providing continuing education for professionals and general education to the public”[2].

In October 2008, NSBSP started to control all screening and diagnostic mammography exams [2]. The vision of the program is to "provide quality standardized mammography access with timely assessment, informed patient navigation and appropriate follow-up of women who have abnormal mammograms on screening, through diagnostic work-ups in accredited work-up centers, before consideration of surgical alternatives” [1]. This program provides breast cancer screening services through eleven health organizations and three mobile vans distributed throughout the province [2].

1. Introduction

Breast cancer is one of the most commonly diagnosed diseases, as well as one of the leading causes of cancer-related deaths, among women. According to the World Health Organization (WHO), breast cancer caused 460,000 deaths world-wide in 2008 and this number is likely to increase within the next few years [3]. Among Canadian women, breast cancer accounts for 28 percent of cancer cases. According to the Canadian Breast Cancer Foundation (2011), "in 2011, an estimated 23,400 women in Canada will be diagnosed with breast cancer, an increase of 200 from 2010. On average, 450 Canadian women will be diagnosed with breast cancer every week" [4].

The purpose of performing the breast screening procedure is to detect breast cancer in the early stages, before symptoms occur (American Cancer Society). Breast screening is an exam that helps to detect breast cancer in people who do not have any symptoms. This approach offers better opportunities to cure the disease [5].

As NSBSP has increased public awareness regarding the importance of breast screening, the program has started to suffer from long wait times. In response to this downside of their success, NSBSP has analyzed wait times by generating reports. However, generating these kinds of reports consumes substantial time and resources. In order to address this issue in a more time- and cost-effective way, health informaticians and statisticians have begun developing tools that help to auto-generate wait-time reports. Such tools will help pinpoint the reasons for the long wait times.

Tools such as these can be developed by numerous statistical analysis techniques. These analysis techniques can be used automatically within a short time instead of working manually for days to find results. In the presence of many free softwares that have the ability to analyze the data statistically, R has become the most popular program, since there are many libraries that are designed by active partners [6]. Also, Latex is the most popular software system that has been used to create good quality documents [7]. In this project, The author has used R and Latex to develop AutoReport Generator software. As well, the author has used C# programming language to create the software interface. This interface will help the user to select any indicator at any site during any time frame.

2. LITERATURE REVIEW

2.1 Breast Anatomy

The human breast is basically made up of glandular and fatty tissue that function to produce milk. It is located between the second and sixth ribs over the pectoralis major. It is anchored onto the muscle with the help of the deep fascia and ligaments, the pectoralis fascia. The fascia is not taught and thus allows for natural motion of the breast. The ligaments run through the breast tissue and the deep fascia. Inferomedially, the breast overlies the rectus abdominis. The nipple overlies the inframammary line, at the level of the 4th rib (figure 1). It is lateral to the mid clavicular line [8].

Muscles underlying the breast include serratus anterior, rectus abdominis fascia and external oblique. Blood supply to the muscles perforates through the parenchyma of the breast while supplying arterial blood to the breast [8]. The blood supply of the breast is mainly derived from the perforators of the internal mammary artery. Other sources of blood supply are arteries to the serratus anterior muscle, the thoracoacromial artery and the lateral thoracic artery. The blood supply is rich and thus provides scope for reduction procedures in which the flaps need to be viable [8].

Sensory innervation to the breast is mainly dermatomal and is derived from the anteromedial and anterolateral branches of the thoracic intercostal nerves T3-T5. Some innervations from the supraclavicular nerves derived from the cervical plexus also occur towards the lateral and upper portions of the breast. The nipple is innervated by the lateral cutaneous branch of T4 [8].

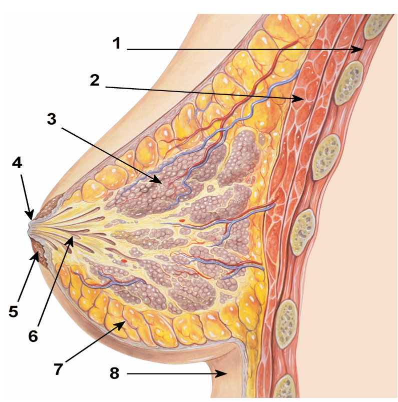

Figure 1: Breast Anatomy; 1. Chest wall, 2. Pectoralis muscles, 3. Lobules, 4. Nipple surface, 5. Areola, 6. Lactiferous duct, 7.Fatty Tissue, 8.Skin [9]

2.2 Breast Cancer

Breast cancer is the most commonly diagnosed cancer among women and is one of the leading causes of cancer related death in them. Cancer in the breast arises because of several molecular alterations at the cellular level, leading to spread and outgrowth of epithelial cells of the breast with uncontrolled growth and immortal features. Among Canadian women, breast cancer accounts for 28 percent of cancer cases. According to Canadian Breast Cancer Foundation (2011), "in 2011, an estimated 23,400 women in Canada will be diagnosed with breast cancer, an increase of 200 from 2010. On average, 450 Canadian women will be diagnosed with breast cancer every week" [10]. The rise in public awareness has improved screening and has facilitated earlier diagnosis, at stages when curative therapies and complete surgical resection is possible. All these have improved the survival rates especially in young women. In this essay, breast cancer and the role of Canadian breast cancer program in the management of breast cancer patients in Canada will be discussed [10].

2.3 Types of Breast Cancer

There are several types of breast cancer and the most commonly diagnosed breast cancer is infiltrating ductal carcinoma. This type of tumor accounts for more than 75 percent of breast cancer cases. It infiltrates the surrounding tissue and also metastasizes via the lymphatics [11]. There are five subtypes of ductal carcinoma and they are cribriform, papillary, comedo, solid and micropapillary. More often than not, the lesions are a combination of 2 or more subtypes. It is important to note that comedonecrosis in a breast tumor is a risk factor for development of ipsilateral breast cancer [11]. Other types of breast cancer include infiltrating lobular carcinoma, lobular carcinoma in situ, medullary carcinoma, tubular carcinoma, papillary carcinoma, metaplastic cancer, mammary Paget disease and colloid carcinoma. In lobular cancer, the cancer develops from the lobules and this occurs in about 10 percent cases [12]. Histologically, the most common type of breast cancer is epithelial tumor [11]. Other histologic types are primary lymphoma, adenocarcinoma and angiosarcoma [12].

2.4 Risk factors

Risk factors for breast cancer development are advanced age, female gender, late age of first parturition, positive family history, atypical hyperplasia lesions in the breast, hormone replacement therapy, lifestyle factors, early menarche,late menopause and genetic predisposition [11]. The most widely recognized risk factor for breast cancer is family history of breast carcinoma in a first-degree relative. According to Swart, "the lifetime risk is up to 4 times higher if a mother and sister are affected; the risk is approximately 5 times greater in women with 2 or more first-degree relatives with breast cancer; and it is also greater among women with breast cancer in a single first-degree relative, particularly if the relative was diagnosed at an early age (50 y or younger)."

2.5 Stages of Breast Cancer

There are five stages of breast cancer [12]. In stage 0 the breast cancer is noninvasive in nature. The carcinoma is in situ and there is no metastases or lymph node involvement. In stage-1 the tumor size is less than 2 cm and the tumor is not spread from the breast. In stage-2, the cancer is spread to ipsilateral axillary lymph nodes. In stage-3, the tumor is more than 5 cm and more lymph nodes are involved. In stage-4 there is metastases [12].

2.6 Prognostic Factors

Prognostic factors related to breast cancer include size of the tumor, status of axillary lymph node, vascular or lymphatic invasion, patient age, histologic grade, response to neoadjuvant therapy, histologic grades and subtypes and amplification or over expression of HER2 gene. Prognosis in stages-0 and 1 is above 95 percent and in stage-4 is 20 percent [11]. Noninvasive lesions are mainly limited by basement membrane.

Though there has been an increase in the number of breast cancer cases worldwide, deaths related to cancer have decreased because of early detection and improved modalities of treatment. Mammogram is a useful screening and diagnostic test for cancer of the breast. Biopsy and magnetic resonance imaging are useful confirmatory tests [13].

2.7 Diagnosis and Treatment

Breast, like any other tissue in the body can be afflicted by pathology and disease, the worst and medically challenging being breast cancer. In fact, breast cancer is the most common cause of cancer-related deaths among women in several parts of the world [14]. Medical imaging has a major role to play in the detection and evaluation of breast cancer and other related pathological problems. The most widely used medical imaging technique is mammography. This is used both for screening and diagnosis of breast tumors [14]. Mammography is an X-ray imaging technique that is specialized and helps in creating images of the breast that appear in detail. The technique employs low doses of X-rays for the purpose. The contrast and resolution are high and help to demonstrate <100-micrometer microcalcifications [14].

Other imaging techniques include magnetic resonance imaging or MRI and computed tomography (CT scan). Both these techniques have adjuvant roles in establishing diagnoses of breast tumors. MRI is more useful for screening and evaluation of dense breasts in women less than 40 years of age, especially in those who are at risk for development of breast cancer. CT scan is useful to monitor the spread of the disease. It is however associated with some risk of radiation and must be performed only in those in whom MRI is contraindicated [14].

The role of ultrasound (US) in breast imaging has also been identified. It is useful to differentiate solid masses from cysts. It is also used to provide guidance in various interventional procedures. US is not useful for screening [14].

Primary treatment for breast cancer is surgery and in those with early stage, the disease is curable. The goals of surgery in breast cancer are "complete resection of the primary tumor with negative margins to reduce the risk of local recurrences, and pathologic staging of the tumor and axillary lymph nodes for providing necessary prognostic information" [13]. Other modes of treatment include chemotherapeutic agents, radiation therapy and biologic agents.

Thus, breast cancer is one of the common causes of cancer related deaths in women. Screening is the best way to decrease mortality and morbidity related to breast cancer. Women need to be educated about the importance of breast cancer screening to facilitate early diagnosis and management.

3. Internship Work

For several years, NSBSP has utilized two systems: the Mammography Information System (MIS), for booking breast screening appointments, and the Diagnostic Reporting System (DRS), for booking breast diagnostic procedures. These systems are relatively old and not very useful for data reuse and analysis. Data from these two systems are stored in one table in the database.

Wait times at NSBSP used to be calculated using a prospective method, meaning they would look forward in the calendar and calculate the time that a patient had to wait if, for instance, she were to call today. However, this method of calculating wait times is misleading and also has caused some issues. NSBSP workers knew when the wait times would be calculated and therefore intentionally used to leave some appointments empty at the end of that month. When the actual calculation took place, the manager would check to see if a patient had called and how long she would have to wait to be booked. Since, as mentioned, some appointments were left empty by clerks, the manager would calculate the wait time from today to the first ‘empty’ appointments.

NSPSP now uses a retrospective method to calculate wait times. This method calculates how long patients had to wait to receive healthcare service or procedures instead of calculating how long a patient would wait to receive health care service or procedure. Using this method of calculating wait times eliminates the biases that were introduced using the old method. Now, wait time reports are calculated using the retrospective method. However, since it is difficult to retrieve data from the old systems, migration to a better system has been introduced.

NSBSP has decided to upgrade their system to a better data management system called CAISIS. CAISIS is an open-source web-based clinical data management system. It is freely distributed from the website www.caisis.org. CAISIS is built using a Microsoft SQL server and C# in .Net platform. It provides virtual and real-time data without relying on any physical location. CAISIS integrates patient care, clinical workflows and research [15]. An example of CAISIS interface can be seen in Appendix A.

CAISIS stores key patient data in chronological order. The data is stored in a relational database that can be called back at each clinic for modifications, and is updated later. The patient histories can then be summarized using computer algorithms and printed on clinic forms. Both clinic and research data are stored logically separated in CAISIS so that clinical decisions are not affected by research information [15][16].

Our work in this internship will involve the use of CAISIS and the retrospective method along with other software such as R and Latex to auto-generate wait time reports.

3.1 Works Performed Step by Step

At the beginning of the internship project, the author had a meeting with the supervisors, Mr. Mohamed Abdolell and Mrs. Theresa Foley, to obtain the instructions and the requirements to accomplish the project. During this meeting, he exposed to the problem that should be solved with the newly developed software. Finally, Mrs. Foley set an appointment to meet with her at the NSBSP office to show and explain to the author how workflow is proceeding in their institution.

3.1.1 Workflow

Before starting the planning process of the project, it is very important to know how successful the current workflow is at NSBSP. The process starts with answering phone call inquiries and receiving paper-based requests sent by fax from family physicians offices, requesting appointments. Appointment clerks then have to check patient information and look for the status of the request to see if it is “Urgent”, “Semi-Urgent” or a routine exam, before assigning and appointment.

3.1.2 The Work Process

At the request of the project supervisor, the author installed Caisis, R statistic software, Latex software (MikTex) and MySQL database management system. After these softwares were downloaded, the author explored how each one could be used.

3.1.2.1 Getting Used to R: A Statistical Analysis System

R is an open-source environment that can be used to apply statistical methods on data. The benefits of R are that it is free, it has an excellent help system, excellent graphic capabilities and can be used as a programming language [17][18]. Moreover, it has the ability to incorporate with open database connectivity (ODBC) by using RODBC package.

In this step, the author executed some codes to create tables and plots in the R system. He then trained himself to access the Caisis database through R and extract data from it with simple SQL queries. At the end of this step, he learned how to create and save a pdf file that contains plots and tables created by R in a specific folder in the computer.

3.1.2.2 Getting Used to Latex: A Document Preparation System

Latex is a software system that is used to produce high quality documents. It is especially good for technical documents and is available for all computer platforms [7]. The Windows version of Latex is called MikTex.

This was an easy language for the author to learn, and he created a style sheet to obtain the standard document design. At the end of the second step, the first prototype of the report to be automatically generated was developed. Figure 2 below shows the style sheet.

Figure 2: Latex template

3.1.2.3 Combining R and Latex Together

In this step, the author worked on combining R with Latex to get the Latex document with plots and tables automatically after running codes in R. At the beginning, he learned how to produce a .tex file using an editor and how to compile it using the Latex Compilers. Latex Compilers will compile an error-free file into .dvi file, whereas the compiler pdflatex produces a pdf file. The output files are to be viewed using appropriate readers, such as xdvi for .dvi and acrobat for pdf. Then, by searching online, the author found that there are some other packages that need to be downloaded into R (e.g., Reporttool and ReportR) to allow R and Latex to function together and create pdf documents automatically. By using these package functions, the intermediary .tex file created by R will pass to MikTex to automatically transform the report into a pdf document format without having to perform any additional steps.

3.1.2.4 Identifying Needed Tables and Attributes

In this step, at the beginning, the author worked on mapping the MIS and DRS database with Caisis database. During the mapping step, the author found some attributes were not will mapped during data migration. Therefore, he needed to define the right way to migrate those attributes. Later on, the author formulated the required equations to obtain the wait times for eight different indicators. Those indicators and their equations to calculate them are as follows:

- Screening wait time: (Date of first time screen - Date of call or requesting).
  The National Target for this waiting time is 90 days.

- Workup wait time: (Date of workup - Date of abnormal screen).
  The National Target for this waiting time is 35 days.

- Screen to Report wait time: (Date of final report - Date of screen).
  The National Target for this waiting time is 14 days.

- Report to workup: (Date of workup - Date of final report).

The National Target for this waiting time is 21 days.

- Diagnostic wait time: (Date of diagnostic mammogram - Date of contact).
  The National Target for this waiting time is 14 days.

- Screen to Core wait time: (Date of core biopsy - Date of abnormal screen).
  The National Target for this waiting time is 49 days.

- Image to Core wait time: (Date of core biopsy - Date of last related imaging procedure).
  The National Target for this waiting time is not available.

- Core to surgery: (Date of Surgery - Date of Core Biopsy).
  The National Target for this waiting time is not available.

After the author had completed formulating these equations, he started writing the SQL queries to retrieve the requisite data from the Caisis database and integrate it with R codes.

3.1.2.5 Using R System to Calculate P90, P50 (Median)

After the author retrieved the needed data for a specific indicator from the Caisis database in R, he grouped this data by year and quarters. Next, the 90th percentile and the median for each group were calculated (i.e., data retrieved for 2009 and 2010 were sectioned into eight groups: Q1/2009, Q2/2009, Q3/2009, Q4/2009, Q1/2010, Q2/2010, Q3/2010 and Q4/2010, after which the P90 and the median for each group were calculated).

3.1.2.6 Plotting the Results from the Previous Step

In order to plot the result in a line chart, the author downloaded the graph packages for R. By using plot functions, the following chart was created. An example of the line chart can be seen in Figure 3.

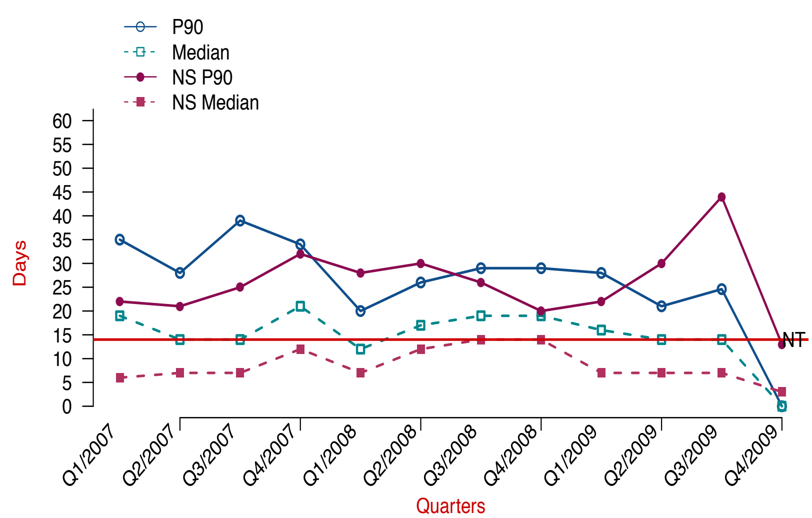

Figure 3: An example of the line chart

Another type of chart created was the bar chart. This chart was needed to compare indicators between all sites. An example of a bar chart is shown below, in Figure 4.

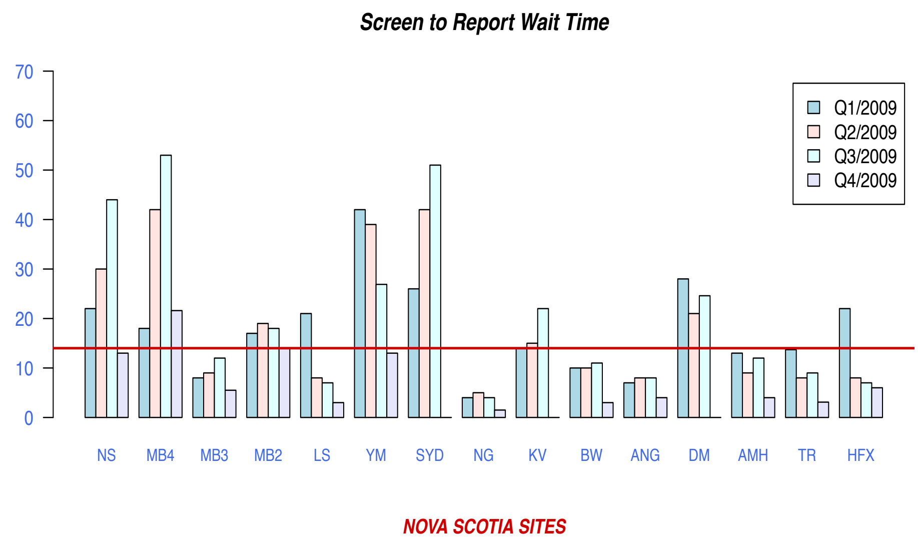

Figure 4: An example of a bar chart

3.1.2.7 Using Statistical Process Control (SPC)

SPC refers to the usage of statistical tools with the intent of establishing control and monitoring a process to ensure that there neither is any lack of potentiality within the process nor any deviation (by any means) among the outputs [19]. Control limits are the lines (a specified distance is maintained from the central line) in the control chart which are calculated by using statistical tools and indicate the behavior of the process i.e., whether the process is 'out of control' or 'in-control'. One of the SPC's most important benefits is that it can provide valuable prediction of the future workflow performance [20]. Also, SPC helps managers to identify the areas that need to be improved. It can detect sudden changes in system performance and allow the manager to react quickly [20].

In order to calculate UCL and LCL, the author downloaded the SPC package in R. An example of a statistical control chart is shown below in Figure 5.

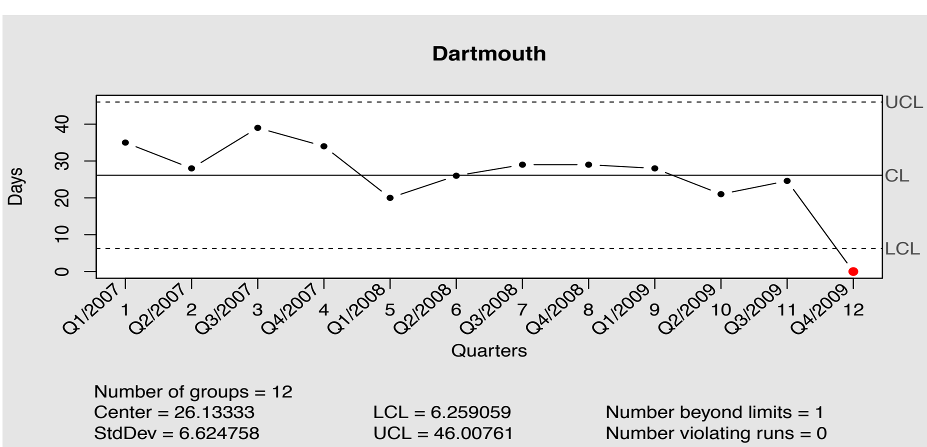

Figure 5: An example of a SPC Chart

3.1.2.8 Creating Software Interface

The author worked on this interface by using Microsoft Visual C# 2010. He learned basic functions from several tutorials offered on the YouTube website. The user interface allows users to choose the indicator and the site. An example of the user interface can be seen in AppendixB.

4. The Final Result

The AutoReport Generator Interface was created with C# language in form type. The author used this language because it would be easier to implement C# codes in CAISIS software in the future. Selected items from the user interface pass to a batch file, which is responsible for running R and Latex software in the background and for passing the chosen items to R codes and SQL queries. At the time of the report's creation, the intermediary .tex file created by R will be passed to MikTex to transform the report into a pdf document format. Finally, the AutoReport Generator will delete all temporary files created during the process. The model of the AutoReport Generator can be seen in Figure 6 and Figure 7.

The Beamer-style sheet file is used to obtain a standard document design that can be opened by the Latex program. Also, there are .R files which can be opened by Tinn-R or Notepad and run by R software. Those files contain R and Latex codes. ODBC connection is used to connect the database with the MySQL Server. All softwares and programming languages used in this project were requested from the project supervisor. According to the supervisor, the R system allows developers to add additional information and functions to existing commands and plots to make them more customized. The R system also allows us to apply statistical process control on our data.

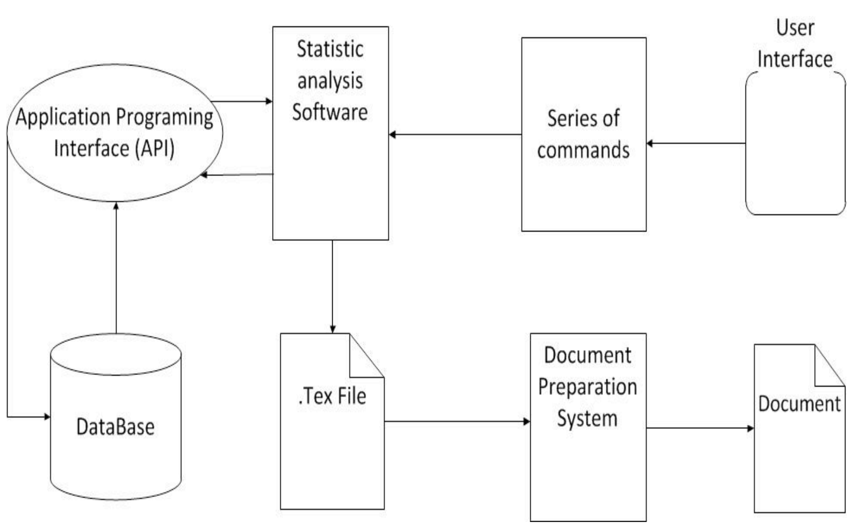

Figure 6: AutoReport Generator software architecture

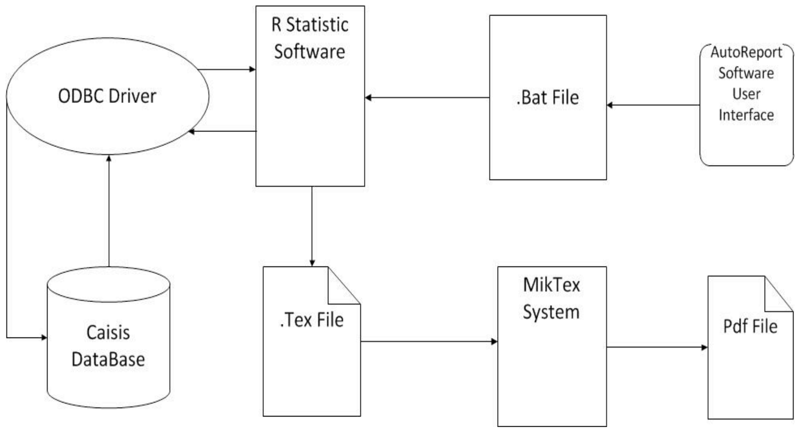

Figure 7: AutoReport Generator software architecture

This software has the ability to calculate P90, Median, UCL and LCL for each quarter of the year for each site. This calculation depends on the wait time indicator and includes eight indicators, as follows:

- Screening wait time
- Workup wait time
- Screen to Report wait time
- Screen to Core wait time
- Report to workup
- Diagnostic wait time
- Image to Core wait time
- Core to surgery

It also includes 11 sites, as follows:

- Amherst
- Antigonish
- Bridgewater
- Dartmouth
- Halifax
- Kentville
- Lower Sackville
- Mobile van 2
- Mobile van 3
- Mobile van 4
- New Glasgow
- Sydney
- Truro
- Yarmouth

An example of the final wait time report generated by the software can be seen in Figure 8.

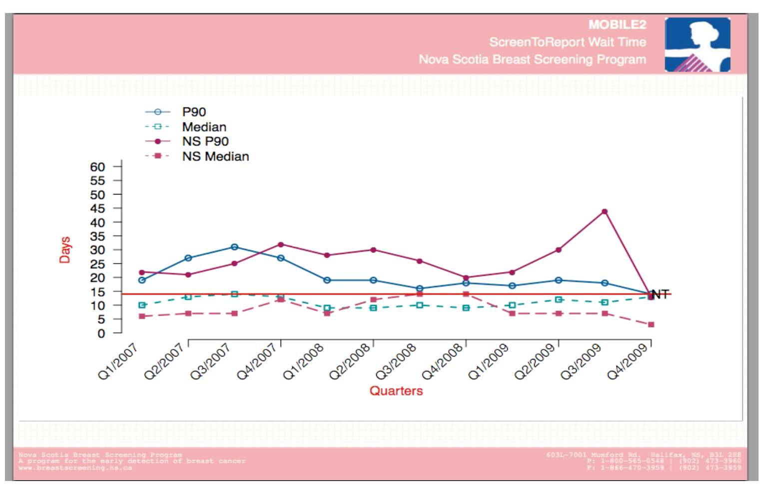

Figure 8: An example of a final report generated by the software

5. Benefits of the Software

This software was designed with the following considerations in mind:

- This software is suited to Breast Screening Program managers who prefer to stay away from command codes and who are interested in automated solutions to obtain excellent documented reports.
- The reports generated will be used to standardize the statistical analysis of the data at all times.
- The reports will save employees time. Instead of taking several days to get the report ready, it will take only a few minutes.

6. Delivery of the Software

After completing the development of the software, a presentation was done on August 10, 2011, to explain the software's features and how it works. Dr. Judy Caines, Medical Director of the NSBSP, Dr. Penny Banres, Pathologist at CDHA, Professor Mohamed Abdolell, Dalhousie facility member, and Mrs. Theresa Foley, Manager of NSBSP as well as internship supervisor, audited this presentation. At the end of the presentation, they expressed their pleasure and excitement over the results. Finally, the author received an accomplishment letter from the supervisor (see Appendix C).

7. The Relationship with Health Informatics

During the internship, the author met with NSBSP staff and reviewed the current processes to obtain solid background information regarding current workflow. This work was strongly associated with the content of the Health Informatics courses that the author took during the Master of Health Informatics program. For instance, the author used the skills he had gained from the "Networks and Web for Health Informatics" course to build the software. Additionally, in this course, the author learned how to map two databases, the structure of the SQL quires, and the basic concepts of programming languages. This latter knowledge acquisition has allowed him to easily learn other new languages (e.g., in his project, the author used C#, R and Latex programming languages). Moreover, in the "Statistics for Health Informatics" course, the author learned how to use statistics to discover the relation between different attributes in the same datasets. In this project, the author used Median and 90th percentile to calculate the number of women who received the service before a given date. The median was chosen because it would not be affected by outliers. Also, the author used upper and lower control limits to monitor whether wait times are statistically 'in-control' or 'out of control'. Furthermore, in the "Project Management for Health Information Projects" course, the author learned how to deal with new people and new projects. Finally, in the "Flow and Standards" course, the author discovered that standardizing work and information flow helps to improve healthcare services.

8. The Author's Role in this Project

The author's background is mainly medical in nature. This posed an issue of miscommunication with information technology (IT) staff, an issue that was reflected on a larger scale in the actual workflow analyzed by the author. In dealing with Mr. Al Qatami who has a more technical background, the author became more adept in proper communications of this nature and attempted to apply the concept on a bigger scale in the actual workflow scenario. The whole process illuminated the issue of communication gaps between medical and IT staff as well as the proper ways to exchange information between them.

To bring to completion, the project required both types of experience, medical and technical. The complementary nature of the author's and Mr. Al Qatami's skill sets proved to be beneficial in this regard. The medical aspects were useful in identifying wait time indicators, workflow analysis and attribute selections. In addition, the technical aspect was useful in programming. Even though each of the two roles fell in the comfort zones of the author and Mr. Alqatami, respectively, they both shared responsibility for completing all tasks as a team.

9. Challenges

Challenges that the author faced during this project include the following:

- Dealing with an incomplete database missing many attributes was a really hard task. The author needed to look at the old database to specify the tables and attributes needed and then map it to the new database that was integrated with CAISIS software. This was a challenge, due to my limited programming skills.
- The database administrator, who was responsible for the migration task, was a busy person. Therefore, any changes that requested and thought were necessary to the database took a long time to make, which greatly delayed the internship process. This also led to an increased workload at the end of the internship duration.
- Learning three new programming languages within the internship period was also a challenging task. The author had to learn the R environment, as well as the Latex and C# programming languages. As previously stated, this was a challenge because of his limited skills base in terms of programming.

10. Conclusion

While working with the NSBSP, the author obtained some valuable experience that ultimately enhanced his own skills. The author applied the knowledge that he gained during pursuing his degree of Health Informatics. The experience that was gained during the internship includes the following:

- Documenting workflow process for the NSBSP was a valuable method to identify the parameters that were used to create the equations for all wait time indicators. Parameters such as date of call, date of screening and date of workup, etc., were easier to identify after understanding the workflow. Also, knowing the degree of urgency helped to exclude the non-urgent cases in some of the wait time indicators, which otherwise would have changed the exact result.

- Being exposed to new softwares offered the author the opportunity to learn three new programing languages and improve his programing skills. These three languages will certainly be used by the author in his professional career.
- Working with NSBSP allowed the author to understand the Canadian breast screening system. This experience can be applied in his home country, where he can establish or improve the breast-screening program there.
- Using open source programs can help to develop tools that have valuable benefits, i.e., the use of the AutoReport Generator will give NSBSP the means to easily perform immediate quality control tasks and improve the performance of the program.

11. Recommendations

It was mentioned that one of goals of implementing CAISIS was to standardize the system that is being used by all Canadian breast screening programs. To do so, the author would like to make the following recommendations:

- ➢ All SQL queries that were written in this internship were specific only to the NSBSP. The AutoReport generator will not work if CAISIS is given to another Canadian breast screening program. The author recommends that NSBSP modify the CAISIS database and import more tables and attributes, such as adding SiteID to the Institutions and Appointments tables so that more generalized queries can be used. As a result, the measurement method of the waiting time will be standardized in all Canadian breast cancer screening programs.
- ➢ Also, the author recommends that NSBSP apply data mining techniques to extract knowledge such as factors associated with the wait times. This will assist the NSBSP manager in finding the reasons behind long wait times and untimely reduce the wait times.

After visiting the NSBSP and reviewing their workflow, the author would like to recommend the following points:

- ➢ An automated booking system, such as online or over-the-phone booking, can be used instead of the current manual system. With this system, all human errors such as leaving empty appointment slots during the bookings can be avoided.

12. References

[1] Nova Scotia Breast Screening Program. (2010). Nova Scotia Breast Screening Program Annual Report 2010. Retrieved August 22, 2011 from: http://www.breastscreening.ns.ca/ann_rpt_2010/

[2] Nova Scotia Breast Screening Program. (2010). Who We Are. Retrieved August 22, 2011 from: http://www.breastscreening.nshealth.ca/who-we-are

[3] Media Centre, World Health Organization. (2011) Fact Sheet, Cancer. Retrieved August 17, 2011 from: http://www.who.int/mediacentre/factsheets/fs297/en/index.html

[4] Canadian Breast Cancer Foundation. (2011). Breast Cancer In Canada 2011. Retrieved August 10, 2011 from: http://www.cbcf.org/enUS/About%20CBCF/Media%20Centre/News%20Releases/2011%20Breast%20Cancer%20Stats.aspx

[5] The Nova Scotia Breast Screening Team. (2009). Program Information, Accountability and Quality. Retrieved August 17, 2011 from http://www.breastscreening.ns.ca/index.php/programinformation

[6] Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004).Bioconductor: open software development for computational biology and bioinformatics. Genome biology, 5(10), R80. doi: 10.1186/gb-2004-5-10-r80.

[7] Lamport Leslie. (2000). Latex A Document Preparation System. Pearson Education, Inc (pp. 11-15)

[7] Gabriel, A. (2011). Breast Anatomy. Medscape Reference. Retrieved August 10th , 2011 from: http://emedicine.medscape.com/article/1273133-overview

[8] Infrared medical Solutions. (2008). Breast Anatomy, retrived 24 August, 2011 from: http://www.infraredmedicalsolutions.com/patients/about_breast_physiology

[9] Canadian Breast Cancer Foundation. (2011). BREAST CANCER IN CANADA 2011. Retrieved August 10th, 2011 from: http://www.cbcf.org/enUS/About%20CBCF/Media%20Centre/News%20Releases/2011%20Bre ast%20Cancer%20Stats.aspx

[10] Wright, M.J., and Moroz, K. (2009). Breast Cancer. Emedicine from WebMD. Retrieved August 10th, 2011 from: http://emedicine.medscape.com/article/1276001-overview

[11] Buechler, M., and Buechler, E. (2007). Breast Cancer. Emedicinehealth. Retrieved August 10th, 2011 from: http://www.emedicinehealth.com/breast_cancer/page14_em.htm#Authors%20and%20Editors

[12] Swart, R. (2011). Breast Cancer. Medscape Reference. Retrieved August 10th, 2011 from: http://emedicine.medscape.com/article/1947145-overview'

[13] Dongola, N. (2011). Mammography in Breast Cancer. Medscape Reference. Retrieved August 10th, 2011 from: http://emedicine.medscape.com/article/346529-overview#a30

[14] Paper, D., Charitable, A., Shortt, F. S. E. D., & Queen, H. P. (2000). Waiting for Medical Services in Ontario : Clarifying the Issues In a Period of Health Reform. Health Policy. (pp. 4-8)

[15] Khushi, M., Carpenter, J. E., Balleine, R. L., & Clarke, C. L. (2011). Development of a data entry auditing protocol and quality assurance for a tissue bank database. Cell and tissue banking. doi: 10.1007/s10561-011-9240-x.

[16] Caisis team. (2002). What can Caisis do for physicians? Project history. Retrieved August 20, 2011from: http://www.caisis.org/forPhysicians.aspx

[17] R Cran team. (2011). The Comprehensive R Archive Network. Retrieved August 15, 2011 from: http://cran.r-project.org/

[18] Venabales D.M., Smithe D.M. and the Core Development Team. (2009), An Introduction to R, R: A Programming Environment for Data Analysis and Graphics. (pp.2-4).

[19] Oakland, J S. (2003). Statistical Process Control. Butterworth-Heinemann, (Fifth edition).

[20] Thor, J., Lundberg, J., Ask, J., Olsson, J., Carli, C., Härenstam, K. P., et al. (2007). Application of statistical process control in healthcare improvement: systematic review. *Quality & safety in health care*, *16*(5), 387-99. doi: 10.1136/qshc.2006.022194. (pp. 393).

APPENDIX A: Caisis software user interface.

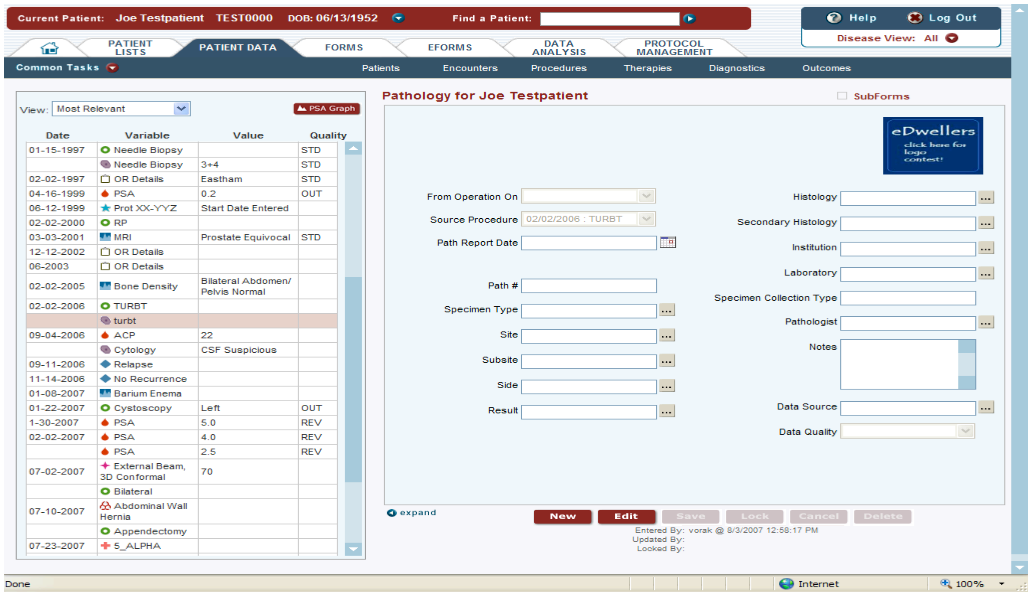

Figure 9: Caisis software user interface.

APPENDIX B: AutoReport Generator Software User Interface

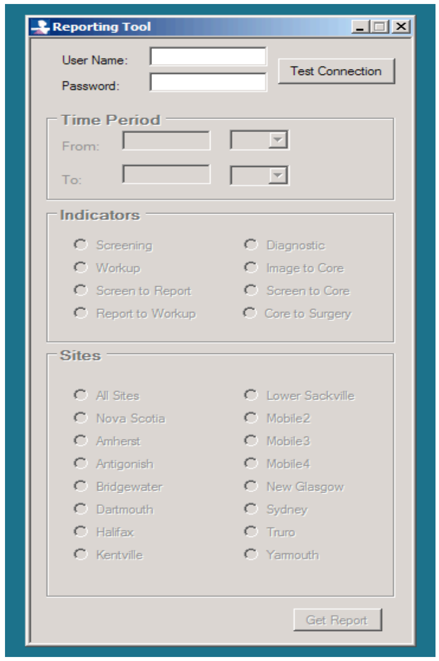

Figure 10: AutoReport Generator login page

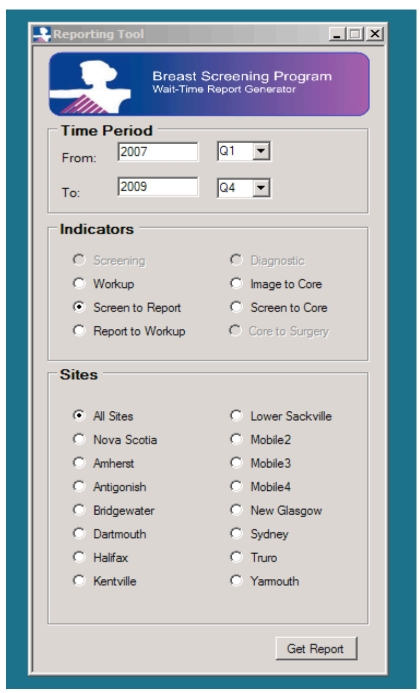

Figure 11: AutoReport Generator User interface

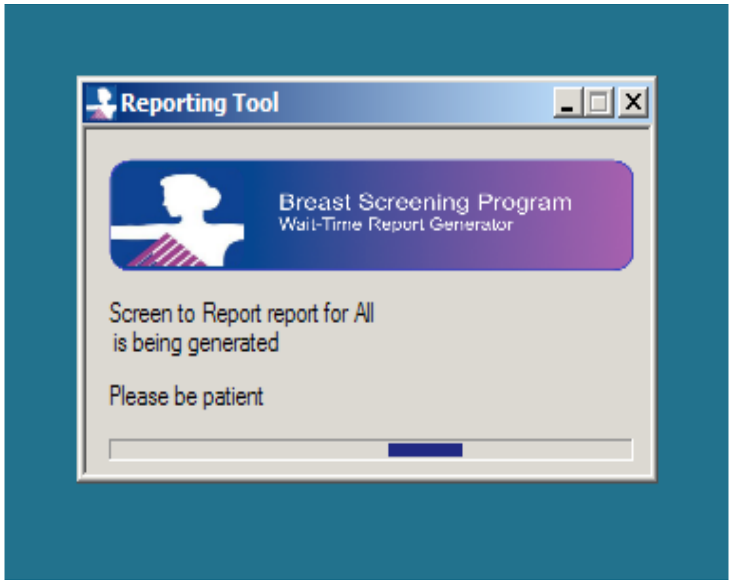

Figure 12: AutoReport Generator processing page

APPENDIX C: Accomplishment Letter

Nova Scotia Breast Screening Program
603L – 7001 Mumford Rd.
Halifax, NS B3L 2H8
P: (902) 473-3960
F: (902) 473-3959
www.breastscreening.nshealth.ca

August 22, 2011

Dr. Grace Paterson, PhD
Assistant Professor, Health Informatics
Room 2L-C2 Tupper Building, Dalhousie University
Halifax, NS, Canada B3H 4H7

Subject: Abdulhady Habash's Internship

Dr. Paterson,

This letter is to indicate that Abdulhady Habash has completed all requirements for his internship with the Nova Scotia Breast Screening Program (NSBSP). Abdulhady's ability to push his knowledge and skill set beyond his academic training during his internship is not something easily achieved by most. His work will become a major part of a valuable tool used on a regular basis at the NSBSP.

If you have any questions or would like to discuss the NSBSP's experience with MHI interns, please do not hesitate to contact me.

Sincerely,

Theresa Foley
Program Manager
Nova Scotia Breast Screening Program
603L - 7001 Mumford Rd
Halifax, NS B3L 2H8
P: (902) 473-3956
F: (902) 473-3959

Glossary & Nomenclature

Adinocarcinoma: Any one of a large group of malignant epithelial cell tumors of the glandular tissue [1].

Angiosarcoma: a malignant neoplasm arising from vascular endothelial cells [2].

Carcinomas: An invasive malignant tumor derived from epithelial tissue that tends to metastasize to other areas of the body [3].

Comedonecrosis: type of necrosis occurring with glands in which there is central luminal inflammation with devitalized cells, usually occurring in the breast in intraductal carcinoma [4].
CT: Computed Tomography Imaging.

Histologic: ertaining to the study of the microscopic anatomic and physiologic characteristics of tissues and the cells found therein [5]**.**
Hyperplasia: abnormal increase in the number of normal cells in normal arrangement in an organ or tissue, which increases its volume [6].

MRI: Magnetic Resonance Imaging.

NSBSP: Nova Scotia Breast Screening Program.

MIS: Mammography Information System DRS. Diagnostic Reporting System
SQL: Structured Query Language RODBC. Open Database Connectivity for R
Median: 50% had the service before this point.

P90: 90% had the service before this point.

SPC: Statistical Process Control.

UCL: Upper Control Limit.

LCL: Lower Control Limit.

US: Ultrasound medical Imaging.

---

1 The American Heritage, Medical Dictionary. (2007), by Houghton Mifflin Company, retrieved August 25, 2011 from: http://medical-dictionary.thefreedictionary.com/Carcinomas

2 Dorland's Medical Dictionary for Health Consumers. (2007) by Saunders, an imprint of Elsevier. Retrieved August 25, 2011 from: http://medical-dictionary.thefreedictionary.com/adenocarcinoma

3 Dorland's Medical Dictionary for Health Consumers. (2007). by Saunders, an imprint of Elsevier. Retrieved August 25, 2011 from: http://medical-dictionary.thefreedictionary.com/Angiosarcoma

4 Mosby's Medical Dictionary, 8th edition. (2009). By Elsevier. Retrieved August 25, 2011 from: **http://medical-dictionary.thefreedictionary.com/Histologic**

5 Dorland's Medical Dictionary for Health Consumers. (2007) by Saunders, an imprint of Elsevier. Retrieved August 25, 2011 from: http://medical-dictionary.thefreedictionary.com/Hyperplasia

6 Stedman's Medical Dictionary. (2006) by Lippincott Williams & Wilkins. Retrieved August 25, 2011 from: **http://www.medilexicon.com/medicaldictionary.php?t=19221**

Declaration

This report has been written by me and has not received any previous academic credit at this or any other institution.

___________________________ ____August 22, 2011____

Abdulhady Habash Date